(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79257 A2

(51) International Patent Classification[7]: C07K

(21) International Application Number: PCT/US01/12230

(22) International Filing Date: 12 April 2001 (12.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/197,349 14 April 2000 (14.04.2000) US

(71) Applicant: **PHYTERA, INC.** [US/US]; 377 Plantation Street, Worcester, MA 01605 (US).

(72) Inventors: **DAVIS, Deborah, vanRiet**; 157 Redstone Hill Road, Sterling, MA 01564 (US). **ROGERS, Bruce, Lee**; 54 Richardson Road, Belmont, MA 02478 (US). **WHITE, Abigail, Coffin**; 315 Czesky Road, Hardwick, MA 01037 (US).

(74) Agent: **CLARK, Paul, T.**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/79257 A2

(54) Title: MDR EFFLUX PUMPS

(57) Abstract: The invention features methods of determining whether a nucleotide sequence encodes an MDR efflux pump, methods for deleting a desired region of DNA in a bacterial cell, methods for determining whether a test substance inhibits the growth or metabolism of cells of a strain of *E. faecalis* bacteria having a disruptive mutation in a gene encoding a MDR efflux pump, methods for determining whether a test substance includes a compound that blocks efflux of an antibacterial agent from a cell, and methods for identifying an inhibitor of an MDR pump.

MDR EFFLUX PUMPS

Background of the Invention

Most bacterial cells express proteins that function as multidrug resistance (MDR) efflux pumps, so named because they confer resistance to a large variety of chemically unrelated agents by extruding them from the cell. MDR efflux pumps have been described in phylogenetically diverse organisms including bacteria (Nikaido, J. Bacteriol. 178:5853-5859, 1996), yeast (Kolaczkowski et al., Microb. Drug Resist. 41:143-158, 1998), and mammals. Many MDR efflux pumps are members of multigene families, the best-described of which being the gene families encoding the ATP Binding Cassette (ABC) membrane proteins, the Multiple Facilitator (MF) superfamily, and the small multidrug resistance (Smr) family (Michaelis and Berkower, Cold Spring Harb. Symp. Quant. Biol. 60:291-307, 1995; Paulsen et al., Microbiol. Rev. 60:575-608, 1996). MDR efflux pumps have been shown to contribute significantly to resistance to antibacterial agents in a number of model organisms as well as important Gram-positive pathogens, including *Staphylococcus* (*S.*) *aureus* (Yoshida et al., J. Bacteriol. 172:6942-6949, 1990; Hsieh et al., Proc. Natl. Acad. Sci. USA 95:6602-6606, 1998), *Enterococcus* (*E.*) *faecalis* (Lynch et al., Antimicrob. Agents Chemother. 41:869-871, 1997), and *Streptococcus* (*S.*) *pneumoniae* (Gill et. al., Antimicrob. Agents Chemother. 43: 187-189, 1999).

The determination of the complete DNA sequence of microbial genomes has progressed markedly over the last few years. Model organisms continue to lead the way with the reported sequence of *Saccharomyces* (*S.*) *cerevisiae* (Goffeau et al., Nature 387:1-105, 1997), *Escherichia* (*E.*) *coli*

(Blattner et al., Science 277: 1453-1474, 1997), *Bacillus* (*B.*) *subtilis* (Kunst et al., Nature 390:249-256, 1997) and *Treponema* (*T.*) *pallidum* (Fraser et al., Science 281: 375-388, 1998). Integration of the genetic and biochemical information obtained over decades of research with the newly obtained raw genomic sequence presents an exciting challenge to biologists and bioinformaticists. Many open reading frames (ORFs) have been identified and catalogued based on homology with known genes, or known folding domains, using algorithms such as FASTA and BLAST (reviewed in Storms, "Genome-wide strategies for studying gene function by using model systems." In: Organization of the Prokaryotic Genome, pp. 347-365. Edited by R.L. Charlebois. ASM Press, Washington, D.C. 1999). Significant efforts have been made to use homology relationships between ORFs to establish hypothetical phylogenetic relationships between organisms as well as genealogy of multigene families within any given organism. Two multigene families that have been examined in detail are those encoding the ABC and MF families of proteins. For example, the complete genome of *B. subtilis* led to the identification of no less than 59 possible ABC transporters (Quentin et al., J. Mol. Biol. 287:467-484, 1999). Likewise, analysis of 5885 predicted ORFs in the complete *S. cerevisiae* genome identified 186 potential MF proteins (Nelissen et al., FEMS Microbiol. Rev. 21:113-134, 1997).

Summary of the Invention

In a first aspect, the invention features a method of determining whether a nucleotide sequence encodes an MDR efflux pump. The method includes the steps of: (a) searching a database of nucleotide sequences for sequences having high identity to a sequence encoding a known MDR efflux pump to generate a first set of candidate sequences comprising sequences that

have high identity to the sequence encoding the known MDR efflux pump; (b) from the first set of candidate sequences, selecting the sequences that include a sequence encoding potential transmembrane domains to generate a second set of candidate sequences; (c) in a bacterial cell, mutating the gene corresponding to one of the candidate sequences; and (d) determining whether the bacterial cell exhibits increased sensitivity to antibacterial agents, wherein the candidate sequence encodes an MDR efflux pump if the bacterial cell exhibits increased sensitivity to the antibacterial agents.

In one embodiment, the sequence encoding the known MDR efflux pump is a nucleotide sequence, and the searching step (a) includes comparing the database of nucleotide sequences with the nucleotide sequence encoding the known MDR efflux pump.

In another embodiment, the sequence is a polypeptide sequence, and the searching step (a) includes comparing nucleotide sequences of the database translated into all six reading frames, to the polypeptide sequence of the known MDR efflux pump.

In a related aspect, the invention features a method of determining whether a polypeptide functions as an MDR efflux pump. This method includes the steps of: (a) searching a database of polypeptide sequences for sequences having high identity to a polypeptide sequence that functions as an MDR efflux pump to generate a first set of candidate sequences; (b) from the first set of candidate sequences, selecting sequences that have potential transmembrane domains to generate a second set of candidate sequences; (c) in a bacterial cell, mutating one of the candidate sequences; and (d) determining whether the bacterial cell exhibits increased sensitivity to antibacterial agents, wherein the candidate polypeptide functions as an MDR efflux pump if the cell exhibits increased sensitivity to an antibacterial agent.

In another aspect, the invention features a method for deleting a desired region of DNA in a bacterial cell. The method includes the steps of: (a) transforming bacterial cells with a vector that includes (i) a first region of at least 30 nucleotides substantially identical to a first region of chromosomal DNA in the bacterial cells; (ii) a second region of at least 30 nucleotides substantially identical to a second region of chromosomal DNA in the bacterial cells; and (iii) a third region encoding a polypeptide that provides resistance to a selection agent (e.g., kanamycin), wherein the first and second regions of chromosomal DNA are on opposite sides of the target region of DNA to be deleted; (b) selecting for bacterial cells in which the vector of step (a) has integrated by a first crossover event into the chromosomal DNA by adding the selection agent to the culture medium; (c) culturing the cells selected for in step (b) in the absence of the selection agent to allow for a second crossover event to occur in at least a subset of the cells, wherein the second crossover event results in the loss of (i) the desired region of DNA and (ii) the region encoding the polypeptide that provides resistance to the selection agent in that subset of cells.

The foregoing method can be repeated one or more times, if desired, to delete additional regions of DNA in the same bacterial cell from which the first region was deleted by repeating steps (a) - (c). Accordingly, the invention also features a method for making a multigene mutant bacterial cell. The method can be performed in any bacterial cell that can undergo homologous recombination, including, for example, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Bacillus anthracis*, *Clostridium tetani*, *Clostridium bolulinum*, *Vibrio cholerae*, *Helicobacter pylori*, *Salmonella typhimurium*, *Shigella dysenteriae*, *Bordetella pertussis*, *Yersinia pestis*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Mycobacterium tuberculosis*, *Corynebacterium*

diptheriae, *Borrelia burdorferi*, *Treponema pallidum*, *Enterococcus faecalis*, *Enterococcus faecium*, and *Streptococcus pneumoniae*.

In yet another aspect, the invention features a method for determining whether a test substance inhibits the growth or metabolism of cells of a strain of *E. faecalis* bacteria having a disruptive mutation in a gene encoding a protein selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68). The method includes the steps of: (a) contacting the cells with a test substance; and (b) determining whether the growth or metabolism of the cells is inhibited. An inhibition of growth or metabolism, relative to the growth or metabolism of cells of the strain of bacteria having the mutation but not contacted with the test substance, identifies the test substance as one having the potential use as an antibacterial agent. The bacterial strain can have disruptive mutations in genes encoding at least two of the proteins listed above. For example, the bacterial strain can have disruptive mutations in genes encoding three, four, five, or six or more of the proteins listed above.

In another aspect, the invention features a substantially pure nucleic acid molecule consisting essentially of *E. faecalis Abc1* (SEQ ID NO: 1), *Abc2* (SEQ ID NO: 3), *Abc3* (SEQ ID NO: 5), *Abc4* (SEQ ID NO: 7), *Abc5* (SEQ ID NO: 9), *Abc6* (SEQ ID NO: 11), *Abc7* (SEQ ID NO: 13), *Abc8* (SEQ ID NO: 15), *Abc9* (SEQ ID NO: 17), *Abc10* (SEQ ID NO: 19), *Abc11* (SEQ ID NO: 21), *Abc12* (SEQ ID NO: 23), *Abc13* (SEQ ID NO: 25), *Abc14* (SEQ ID NO: 27), *Abc15* (SEQ ID NO: 29), *Abc16* (SEQ ID NO: 31), *Abc17* (SEQ ID NO: 33), *Abc18* (SEQ ID NO: 35), *Abc19* (SEQ ID NO: 37), *Abc20* (SEQ ID NO: 39), *Abc21* (SEQ ID NO: 41), *Abc22* (SEQ ID NO: 43), *Abc23* (SEQ ID NO: 45), *Bmr* (SEQ ID NO: 47), *NorA* (SEQ ID NO: 49), *Mf1* (SEQ ID NO: 51), *Mf2* (SEQ ID NO: 53), *Mf3* (SEQ ID NO: 55), *Mf4* (SEQ ID NO: 57), *Mf5* (SEQ ID NO: 59), *Mf6* (SEQ ID NO: 61), *Mf7* (SEQ ID NO: 63), *Mate1* (SEQ ID NO: 65), or *Smr1* (SEQ ID NO: 67).

In another aspect, the invention features a substantially pure polypeptide that includes a sequence selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

In still another aspect, the invention features a strain of *E. faecalis* bacteria having a disruptive mutation in a gene encoding a protein selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68). The bacterial strain can have disruptive mutations in genes encoding at least two of the proteins listed above. More preferably, the bacterial strain can have disruptive mutations in genes encoding three, four, five, or six or more of the proteins listed above.

The invention also features a method for determining whether a test substance includes a compound that blocks efflux of an antibacterial agent from a cell. The method includes the steps of: (a) providing a polypeptide selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID

NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68); (b) contacting the polypeptide with the test substance; and (c) determining whether a compound in the test substance binds to the polypeptide, wherein binding of the compound to the polypeptide identifies it as a compound that blocks efflux of an antibacterial agent from a cell. The method may be performed in a cell (e.g., an *E. faecalis* cell, or a heterologous cell) or, alternatively, may be performed in a cell-free system.

In another aspect, the invention features a method for determining whether a test substance decreases expression of an MDR efflux pump selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68). The method includes the steps of: (a) providing a cell expressing the MDR efflux pump; (b) contacting the cell with the test substance; and (c) measuring expression of the MDR efflux pump in

the cell, wherein decreased polypeptide expression, relative to a cell not contacted with the test substance, indicates that the test substance decreases expression of an MDR efflux pump. Expression of the MDR efflux pump can be determined by measuring protein levels of the MDR efflux pump or by measuring levels of RNA encoding the MDR efflux pump. The MDR efflux pump can be expressed in any cell (e.g., a *Bacillus* (*B.*) *subtilis* or *Lactococcus* (*L.*) *lactis* cell). The foregoing method may also be performed in a cell-free system that supports transcription (e.g., whole-cell lysates).

The invention also features a method for identifying an inhibitor of an MDR pump that includes the steps of: (a) expressing an MDR efflux pump derived from a first bacterial strain in a second bacterial strain to provide increased resistance to an antibacterial agent relative to the second bacterial strain not expressing the MDR efflux pump; (b) contacting the second bacterial strain expressing the MDR efflux pump with an amount of the antibacterial agent that inhibits growth of a control strain (i.e., the second bacterial strain not expressing the MDR efflux pump) but does not substantially inhibit growth of the second bacterial strain expressing the MDR efflux pump; (c) contacting the second bacterial strain expressing the MDR efflux pump with a test substance; and (d) measuring growth of second bacterial strain expressing the MDR efflux pump, wherein decreased growth of the bacterial strain identifies the test substance as one that includes an inhibitor of an MDR efflux pump. Step (c) can be performed before step (b), after step (b), or simultaneous with step (b).

The second bacterial strain can be, for example, a strain of *B. subtilis* or *L. lactis*. The MDR efflux pump can be selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO:

12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

The invention also features a method for increasing the sensitivity of a bacterial cell to an antibacterial agent, the method including the step of contacting the cell with a compound that blocks efflux of the antibacterial agent from the cell by binding to an MDR efflux pump selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68). The cell can be an *E. faecalis* bacterial cell, but can also be, for example, any bacterial cell described herein.

In a related aspect, the invention features a method for increasing the sensitivity of a bacterial cell to an antibacterial agent, the method including the step of contacting the cell with a compound that blocks efflux of the antibacterial agent from the cell by decreasing the expression of a gene encoding an MDR efflux pump selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68). The cell can be an *E. faecalis* bacterial cell, but can also be, for example, any bacterial cell described herein.

In still another aspect, the invention features a method for determining whether a test substance includes a compound that reduces efflux of antibacterial agents from a cell. The method includes the steps of: (a) providing a non-bacterial cell expressing a nucleic acid molecule encoding an MDR efflux pump selected from *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID

NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68); (b) contacting the cell with the test substance; (c) contacting the cell with a compound that is capable of (i) entering the cell and (ii) being transported by the MDR efflux pump; and (d) measuring efflux of the compound from the cell, wherein decreased efflux, relative to a cell not contacted with the test substance, identifies the test substance as containing a compound that reduces efflux of antibacterial agents from a cell. The cell can be, for example, a eukaryotic cell, such as a mammalian cell, an insect cell, or a yeast cell (a *Saccharomyces cerevisiae* cell).

The invention also features a method for determining whether a test substance includes a compound that reduces efflux of antibacterial agents from a cell. This method includes the steps of: (a) providing a cell free system consisting of: (i) a lipid membrane into which is inserted an MDR efflux pump selected from *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ

ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68); (ii) the test substance; and (iii) a compound that is capable of being transported by the MDR efflux pump; and (b) measuring transport of the compound across the lipid membrane, wherein decreased transport, relative to transport in a cell-free system not comprising the test substance, identifies the test substance as containing a compound that reduces efflux of antibacterial agents from a cell. In one example, the lipid membrane is in the form of a liposome.

By "MDR efflux pump" is meant any cell transporter that has been functionally shown to actively transport a drug or chemical out of a cell.

By "substantially identical" is meant that 30 or more consecutive nucleotides exhibit at least 50%, preferably 85%, more preferably 90%, and most preferably 95% identity to a reference nucleotide sequence. For homologous recombination, the length of substantial identity will generally be at least 30 nucleotides, preferably at least 50 nucleotides, more preferably at least 100 nucleotides, and most preferably 200 nucleotides.

By "high identity" is meant that two sequences share at least 10% of nucleotides or 20% of amino acids, when optimally aligned, such as by the program BLAST, over a comparison window of 90 nucleotides or 30 amino acids. Preferably, the percent identity is at least 25%, 30%, 40%, or even 50% and more preferably, the identity is at least 70% or even 80%. One sequence may include additions or deletions (i.e., gaps) of 20% or less when compared to the second sequence. Optimal alignment of sequences may be conducted, for example, by the methods of Gish and States (Nature Genet. 3:266-272, 1993), Altshul et al. (J. Mol. Biol. 215:403-410, 1990), Madden et al. (Meth. Enzymol. 266:131-141, 1996), Althsul et al (Nucleic Acids Res. 25:3389-3402, 1997), or Zhang et al (Genome Res. 7:649-656, 1997).

By "disruptive mutation" is meant an alteration in the nucleic acid molecule that results in decreased expression or function of the encoded protein. Preferably, the decrease is at least 50%, more preferably it is at least 80%, and most preferably the decrease is >99%. The mutation can be a point mutation but preferably is an insertional mutation or a deletion mutation. Mutations in the coding region and in the region regulating expression (e.g., the promoter) are each likely to be disruptive.

By "test substance" is meant a compound (or collection of compounds) being tested for its ability to inhibit growth or metabolism of a bacterial strain. Specifically excluded are compounds that were previously known in the public domain to inhibit the growth or metabolism of that species of bacteria. For example, erythromycin would not be considered to be a test substance when administered to a culture of *E. faecalis* because it was known that erythromycin inhibited growth or metabolism of that bacterium.

Any test substance can be used in the present screening methods, including naturally occurring substances and non-naturally occurring substances. Exemplary test substances are low molecular weight substances produced by living organisms or other substances, soluble in the growth medium, having a molecular weight between 150 and 750 daltons. Test substances can be individual compounds or libraries of compounds. One skilled in the art will recognize that there are numerous sources for the test substance, and any of these is suitable for use in the present screening method.

Determining whether the growth or metabolism of a cell is inhibited can be performed using any of a number of standard assays, such as those described herein. Cellular metabolism can be measured, for example, by assessing the activity of one or more enzymes that function in cellular metabolic pathways. Alternatively, cellular metabolism can be measured by any of the surrogate markers that are known to cell biologists. Methods for

determining growth or metabolism include physical observation (e.g., MIC determination) and measuring turbidity (e.g., absorbance at 650 nm; IC_{50} determination).

The method for determining growth or metabolism can be adapted for high throughput screening, for example, by culturing the cells in multi-well or microtiter plates and measuring growth or metabolism spectrophotometrically.

In the absence of any test substance, a bacterial strain having a mutation may exhibit a different rate of growth or metabolism than that exhibited by the parental strain not having the mutation. Accordingly, it is desirable to determine whether a test substance inhibits the growth or metabolism of a cell by comparing the growth or metabolism of a bacterial strain in the presence of the test substance with the growth or metabolism of that strain in the absence of the test substance.

A test substance is considered to inhibit the growth or metabolism of a bacterium if it decreases growth or metabolism by at least 10% when administered at a final concentration of up to 100 μg/ml to a culture of that bacterium, as determined using one of the foregoing assays. Preferably, growth or metabolism is decreased by at least 25%, more preferably by at least 50%, and most preferably by at least 75%.

The invention provides methods and reagents for (i) identifying MDR efflux pumps (ii) knocking out the function of one or more genes encoding MDR efflux pumps in *E. faecalis*, (iii) using these mutated bacterial strains to identify novel antibacterial agents, and (iv) using the novel genes encoding MDR efflux pumps to identify compounds that increase the sensitivity of a bacterial cell to an antibacterial agent. The gene disruption techniques are broadly applicable for knocking out genes in a wide variety of bacteria, including, for example, *Staphylococcus aureus*, *Streptococcus*

pyogenes, *Bacillus anthracis*, *Clostridium tetani*, *Clostridium bolulinum*, *Vibrio cholerae*, *Helicobacter pylori*, *Salmonella typhimurium*, *Shigella dysenteriae*, *Bordetella pertussis*, *Yersinia pestis*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Mycobacterium tuberculosis*, *Corynebacterium diptheriae*, *Borrelia burdorferi*, *Treponema pallidum*, *Enterococcus faecalis*, *Enterococcus faecium*, and *Streptococcus pneumoniae*.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Brief Description of the Drawings

Fig. 1 is a schematic illustration showing single gene knockout generation by gene disruption. Bacterial cells in which the target gene has been disrupted are isolated by selection in the presence of kanamycin. This method cannot be used for the generation of multigene knockout strains because once the kanamycin-bearing plasmid is inserted, a second kanamycin-bearing plasmid can no longer be used for selection of another gene disruption. Additionally, the plasmid backbone integrated into the genome provides an area for non-targeted homologous recombination.

Figs. 2A-2F are a series of schematic illustrations showing *E. faecalis* MDR-knockout strains are more sensitive to antibacterial agents.

Fig. 3 is a schematic illustration showing gene knockout by disruption, followed by marker and gene target deletion. Sequences A and B, which flank the target gene, are amplified by PCR and cloned into a vector that provides kanamycin resistance. The bacteria are transformed with this vector and selected for kanamycin-resistant bacteria (thus selecting for bacteria in which the vector has integrated into the chromosomal DNA due to a first crossover). Subsequently, the cells are grown in the absence of kanamycin.

During this phase, if a second crossover occurs, the target gene and the kanamycin resistance gene are each lost. Thus, the desired bacteria will die in the presence of kanamycin. These kanamycin-sensitive cells can be identified by replicate plating, as described herein.

Fig. 4 is a schematic illustration showing *E. faecalis Abc1* nucleotide sequence (SEQ ID NO: 1) and Abc1 (SEQ ID NO: 2) amino acid sequence.

Fig. 5 is a schematic illustration showing *E. faecalis Abc2* nucleotide sequence (SEQ ID NO: 3) and Abc2 amino acid sequence (SEQ ID NO: 4).

Fig. 6 is a schematic illustration showing *E. faecalis Abc3* nucleotide sequence (SEQ ID NO: 5) and Abc3 amino acid sequence (SEQ ID NO: 6),

Fig. 7 is a schematic illustration showing *E. faecalis Abc4* nucleotide sequence (SEQ ID NO: 7) and Abc4 amino acid sequence (SEQ ID NO: 8).

Fig. 8 is a schematic illustration showing *E. faecalis Abc5* nucleotide sequence (SEQ ID NO: 9) and Abc5 amino acid sequence (SEQ ID NO: 10).

Fig. 9 is a schematic illustration showing *E. faecalis Abc6* nucleotide sequence (SEQ ID NO: 11) and Abc6 amino acid sequence (SEQ ID NO: 12).

Fig. 10 is a schematic illustration showing *E. faecalis Abc7* nucleotide sequence (SEQ ID NO: 13) and Abc7 amino acid sequence (SEQ ID NO: 14),

Fig. 11 is a schematic illustration showing *E. faecalis Abc8* nucleotide sequence (SEQ ID NO: 15) and Abc8 amino acid sequence (SEQ ID NO: 16),

Fig. 12 is a schematic illustration showing *E. faecalis Abc9* nucleotide sequence (SEQ ID NO: 17) and Abc9 amino acid sequence (SEQ ID NO: 18),

Fig. 13 is a schematic illustration showing *E. faecalis Abc10* nucleotide sequence (SEQ ID NO: 19) and Abc10 amino acid sequence (SEQ ID NO: 20),

Fig. 14 is a schematic illustration showing *E. faecalis Abc11* nucleotide sequence (SEQ ID NO: 21) and Abc11 amino acid sequence (SEQ ID NO: 22),

Fig. 15 is a schematic illustration showing *E. faecalis Abc12* nucleotide sequence (SEQ ID NO: 23) and Abc12 amino acid sequence (SEQ ID NO: 24),

Fig. 16 is a schematic illustration showing *E. faecalis Abc13* nucleotide sequence (SEQ ID NO: 25) and Abc13 amino acid sequence (SEQ ID NO: 26),

Fig. 17 is a schematic illustration showing *E. faecalis Abc14* nucleotide sequence (SEQ ID NO: 27) and Abc14 amino acid sequence (SEQ ID NO: 28),

Fig. 18 is a schematic illustration showing *E. faecalis Abc15* nucleotide sequence (SEQ ID NO: 29) and Abc15 amino acid sequence (SEQ ID NO: 30),

Fig. 19 is a schematic illustration showing *E. faecalis Abc16* nucleotide sequence (SEQ ID NO: 31) and Abc16 amino acid sequence (SEQ ID NO: 32),

Fig. 20 is a schematic illustration showing *E. faecalis Abc17* nucleotide sequence (SEQ ID NO: 33) and Abc17 amino acid sequence (SEQ ID NO: 34),

Fig. 21 is a schematic illustration showing *E. faecalis Abc18* nucleotide sequence (SEQ ID NO: 35) and Abc18 amino acid sequence (SEQ ID NO: 36),

Fig. 22 is a schematic illustration showing *E. faecalis Abc19* nucleotide sequence (SEQ ID NO: 37) and Abc19 amino acid sequence (SEQ ID NO: 38),

Fig. 23 is a schematic illustration showing *E. faecalis Abc20* nucleotide sequence (SEQ ID NO: 39) and Abc20 amino acid sequence (SEQ ID NO: 40),

Fig. 24 is a schematic illustration showing *E. faecalis Abc21* nucleotide sequence (SEQ ID NO: 41) and Abc21 amino acid sequence (SEQ ID NO: 42),

Fig. 25 is a schematic illustration showing *E. faecalis Abc22* nucleotide sequence (SEQ ID NO: 43) and Abc22 amino acid sequence (SEQ ID NO: 44),

Fig. 26 is a schematic illustration showing *E. faecalis Abc23* nucleotide sequence (SEQ ID NO: 45) and Abc23 amino acid sequence (SEQ ID NO: 46),

Fig. 27 is a schematic illustration showing *E. faecalis Bmr* nucleotide sequence (SEQ ID NO: 47) and Bmr amino acid sequence (SEQ ID NO: 48),

Fig. 28 is a schematic illustration showing *E. faecalis NorA* nucleotide sequence (SEQ ID NO: 49) and NorA amino acid sequence (SEQ ID NO: 50),

Fig. 29 is a schematic illustration showing *E. faecalis Mf1* nucleotide sequence (SEQ ID NO: 51) and Mf1 amino acid sequence (SEQ ID NO: 52),

Fig. 30 is a schematic illustration showing *E. faecalis Mf2* nucleotide sequence (SEQ ID NO: 53) and Mf2 amino acid sequence (SEQ ID NO: 54),

Fig. 31 is a schematic illustration showing *E. faecalis Mf3* nucleotide sequence (SEQ ID NO: 55) and Mf3 amino acid sequence (SEQ ID NO: 56),

Fig. 32 is a schematic illustration showing *E. faecalis Mf4* nucleotide sequence (SEQ ID NO: 57) and Mf4 amino acid sequence (SEQ ID NO: 58),

Fig. 33 is a schematic illustration showing *E. faecalis Mf5* nucleotide sequence (SEQ ID NO: 59) and Mf5 amino acid sequence (SEQ ID NO: 60),

Fig. 34 is a schematic illustration showing *E. faecalis Mf6* nucleotide sequence (SEQ ID NO: 61) and Mf6 amino acid sequence (SEQ ID NO: 62),

Fig. 35 is a schematic illustration showing *E. faecalis Mf7* nucleotide sequence (SEQ ID NO: 63) and Mf7 amino acid sequence (SEQ ID NO: 64),

Fig. 36 is a schematic illustration showing *E. faecalis Mate1* nucleotide sequence (SEQ ID NO: 65) and Mate1 amino acid sequence (SEQ ID NO: 66),

Fig. 37 is a schematic illustration showing *E. faecalis Smr1* nucleotide sequence (SEQ ID NO: 67) and Smr1 amino acid sequence (SEQ ID NO: 68).

Detailed Description

We have invented a method for identifying MDR efflux pumps by first mining bacterial genome databases to identify candidate pumps, and then validating function using a disruption-mediated gene knockout strategy. Using this method, we have discovered over thirty candidate MDR efflux pumps in *E. faecalis*. Gene knockout experiments to date have verified that at least five of these efflux pumps do function as MDR efflux pumps.

We have also discovered methods for generating bacterial strains in which two or more MDR efflux pumps have been mutated by gene deletion. These bacterial strains are useful for screening test substances for their ability to inhibit the growth or metabolism of bacterial cells.

Example 1: Genomic Sequence Database Mining to Identify Candidate Drug-Efflux Pumps in *E. faecalis*

We used the DNA sequence and membrane topology of several known MDR efflux pumps to probe, using a series of BLAST searches, for homologues within the genomic sequence of *E. faecalis* (Worley et al., Genome Res. 5:173-184, 1995; see Table 1; the database can be accessed at http://www.tigr.org). The amino acid sequence from a known MDR efflux pump from other bacterial and fungal species were used as probes in BLAST searches of the TIGR database (WU-BLAST version 2.0). A list of contigs and their P-values was obtained. P-values under the cut-off of 1e-20 were considered for further analysis. The ORFs from each contig were located and examined for the presence of predicted transmembrane coding regions using Sosui version 1.0/10 (Hirokawa et al., Bioinformatics 14:378-379, 1998). For the ABC family, ABC signature sequences and ATP/GTP binding site motif A (P-loop) using Motif finder (http://www.motif.genome.ad.jp/). ORFs that satisfied the foregoing criteria were then used as probes in BLAST searches against protein databases in order to identify ORFs having orthologues that are known MDR efflux pumps (it is not necessary that an ORF have an orthologue that is a known MDR efflux pump; this step aids in prioritizing candidate MDR efflux pumps for testing).

Table 1

Organism	MDR Pump Gene	Transporter Family	Accession Number
S. aureus	*NorA*	MF	P21191
B. subtilis	*Bmr*	MF	P33449
B. subtilis	*Blt*	MF	L32599
S. pneumoniae	*PmrA*	MF	AJ07367
L. lactis	*LmrP*	MF	X89779
L. lactis	*LmrA*	ABC	U63741
S. epidermidis	*MsrA*	ABC	X52085

The ABC Family

The members of the ABC family of proteins share an ATP binding cassette sequence motif that enables the identification of *E. faecalis* sequences having high identity on the amino acid level to *L. lactis* LmrA and *S. epidermidis* MsrA (designated Abc1 through Abc23, respectively; Table 2). Only a minority of ABC proteins are predicted to act as drug transporters, and, at present, there is no easy way to distinguish MDR-ABC proteins from those that perform any of the myriad of non-MDR related functions within the cell. Thus, a second step of the method includes functional testing of the candidate encoded polypeptides by generating disruption or deletion mutants (see Examples 2 and 4).

The Multiple Facilitator Superfamily

We scanned the *E. faecalis* genomic database using as probes a set of known MDR genes of the MF superfamily from a set of Gram-positive bacteria (e.g. *NorA*, *Bmr*, *Blt*, *PmrA*; Table 1). We identified a set of sequences having high identity to one or more of the probes; these sequences were designated Mf1 through Mf7 (Table 2). The *E. faecalis* genes designated *NorA* and *Bmr* were labeled such because of the high sequence identity to the

equivalent genes in *S. aureus* and *B. subtilis*, respectively (Table 1). Other MF genes in *E. faecalis* are also highly similar to these genes. As with the ABC homologues, the MF family of proteins most likely have a broad range of functions, many of which are unrelated to MDR. Therefore, functional testing by drug sensitivity is performed to definitively designate these proteins as MDR efflux pumps (see below).

MATE and Smr Families

A family of transporters termed MATE (multidrug and toxic compound extrusion transporters) has been described based on sequence homology to the *NorM* multidrug efflux protein of *V. parahaemolyticus* and the *YdhE* gene product of *E. coli* and *Haemophilus* (*H.*) *influenzae* (Morita et al., Antimicrob. Agents Chemother. 42:1778-1782, 1998; Brown et al., Mol. Microbiol. 31:394-395, 1999). The Smr family consists of proteins with four transmembrane-spanning regions, and the MDR efflux proteins within this family are the smallest known secondary transporters (Paulsen et al., Mol. Microbiol.19:1167-1175, 1996). Using the *E. faecalis* genomic database, we scanned for potential members of the MATE transporter protein family with *B. subtilis* Ypn, YisQ, and YucE, and *H. influenza* YdhE, and for potential members of the Smr transporter protein family with *B. subtilis* EbrA and EbrB.

By conducting orthologue searching of MDR efflux pump gene families, we identified candidates for gene-knockout generation and drug sensitivity analysis in *E. faecalis* (Table 2).

Table 2

E. faecalis sequence identifier*	Homologue	Organism	Accession #	E value[a]
Abc1 (SEQ ID NO: 2) (Contig 6453)	Tap-tub2	*M. tuberculosis*	Q11047	9.20E-114
	YfiC	*B. subtilis*	P54719	4.80E-110
	mdl	*E. coli*	P30751	2.50E-107
	Tap-tub	*M. tuberculosis*	Q11046	3.50E-97
	AtrD	*A. nidulans*	AF071411	1.80E-90
	afuMDR1	*A. fumigatus*	U62934	3.30E-90
Abc2 (SEQ ID NO: 4) (Contig 6216)	Tap-tub2	*M. tuberculosis*	Q11047	4.70E-123
	mdl	*E. coli*	P30751	5.30E-117
	YfiC	*B. subtilis*	P54719	6.90E-114
	AtrD	*A. nidulans*	AF071411	6.90E-106
	YfiB	*B. subtilis*	P54718	9.30E-103
Abc3 (SEQ ID NO: 6) (Contig 6222)	YheI	*B. subtilis*	CAB12811	8.40E-106
	ComA	*S. pneumoniae*	M36180	6.30E-57
	afuMDR1	*A. fumigatus*	U62934	1.50E-56
Abc4 (SEQ ID NO: 8) (Contig 6207) Abc5 (SEQ ID NO: 10) (Contig 6207)	Tap-tub2	*M. tuberculosis*	Q11047	1.70E-143
	YfiC	*B. subtilis*	P54719	2.80E-117
	Tap-tub	*M. tuberculosis*	Q11046	1.80E-115
	mdl	*E. coli*	P30751	1.40E-104
	AtrD	*A. nidulans*	AF071411	4.50E-103
	YfiB	*B. subtilis*	P54718	1.70E-101
	afuMDR1	*A. fumigatus*	U62934	4.90E-96
	LmrA	*L. lactis*	U63741	5.20E-73
	ComA	*S. pneumoniae*	M36180	8.40E-58
	pfmdr1	*P. falciparum*	M29154	2.80E-56
Abc6 (SEQ ID NO: 12) (Contig 6344)	LmrA	*L. lactis*	U63741	6.00E-75
	ComA	*S. pneumoniae*	M36180	8.00E-63
	afuMDR1	*A. fumigatus*	U62934	5.60E-56
Abc7 (SEQ ID NO: 14) (Contig 6226)	AtrD	*A. nidulans*	AF071411	5.30E-106
	YfiC	*B. subtilis*	P54719	6.90E-105
	afuMDR1	*A. fumigatus*	U62934	4.70E-103
	YfiB	*B. subtilis*	P54718	6.80E-100
	LmrA	*L. lactis*	U63741	1.70E-71
	ComA	*S. pneumoniae*	M36180	1.40E-57
Abc8 (SEQ ID NO: 16) (Contig 6358)	mdl	*E. coli*	P30751	4.50E-134
	AtrD	*A. nidulans*	AF071411	8.20E-115
	YfiC	*B. subtilis*	P54719	5.80E-108
	afuMDR1	*A. fumigatus*	U62934	1.20E-102
	YheI	*B. subtilis*	CAB12811	4.90E-90
	LmrA	*L. lactis*	U63741	3.10E-71
Abc9 (SEQ ID NO: 18) (Contig 6217)	SunT	*B. subtilis*	CAB14065	7.70E-58

E. faecalis sequence identifier*	Homologue	Organism	Accession #	E value[a]
Abc10 (SEQ ID NO: 20) (Contig 6328)	cydC	*H. influenza*	AAC22811	2.40E-51
Abc11 (SEQ ID NO: 22) (Contig 6185)	Tap-tub2 mdl AtrD Tap-tub afuMDR1	*M. tuberculosis* *E. coli* *A. nidulans* *M. tuberculosis* *A. fumigatus*	Q11047 P30751 AF071411 Q11046 U62934	5.60E-109 1.30E-97 6.10E-97 2.30E-87 1.60E-85
Abc12 (SEQ ID NO: 24) (Contig 6317)	afuMDR1 AtrD pfmdr1	*A. fumigatus* *A. nidulans* *P. falciparum*	U62934 AF071411 M29154	3.30E-81 5.90E-81 4.30E-66
Abc13 (SEQ ID NO: 26) (Contig 6285)	SunT	*B. subtilis*	CAB14065	6.30E-21
Abc14 (SEQ ID NO: 28) (Contig 6864)	mdl	*E. coli*	P30751	7.20E-92
Abc15 (SEQ ID NO: 30) (Contig 6183)	StpC DrrA	*S. aureus* *S. peucetius*	Z30588 P32010	4.60E-27 1.00E-25
Abc16 (SEQ ID NO: 32) (Contig 6229)	MsrA SmrB	*S. epidermidis* *S. ambofaciens*	X52085 X63451	3.10E-88 8.20E-31
Abc17 (SEQ ID NO: 34) (Contig 6174)	SmrB TlrC	*S. ambofaciens* *S. fradiae*	X63451 M57437	7.10E-62 8.00E-57
Abc18 (SEQ ID NO: 36) (Contig 6381)	SmrB TlrC	*S. ambofaciens* *S. fradiae*	X63451 M57437	5.90E-62 2.40E-51
Abc19 (SEQ ID NO: 38) (Contig 6402)	TlrC SmrB	*S. fradiae* *S. ambofaciens*	M57437 X63451	1.00E-46 1.80E-43
Abc20 (SEQ ID NO: 40) (Contig 6203)	DrrA	*S. peucetius*	P32010	2.50E-35
Abc21 (SEQ ID NO: 42) (Contig 6237)	DrrA	*S. peucetius*	P32010	3.40E-24
Abc22 (SEQ ID NO: 44) (Contig 6232)	SmrB TlrC	*S. ambofaciens* *S. fradiae*	X63451 M57437	5.70E-43 5.00E-40
Abc23 (SEQ ID NO: 46) (Contig 6274)	MsrA	*S. epidermidis*	X52085	3.60E-39
Bmr (SEQ ID NO: 48) (Contig 6187)	PmrA	*S. pneumoniae*	AJ007367	7.30E-104
NorA (SEQ ID NO: 50) (Contig 6353)	NorA Bmr Blt	*S. aureus* *B. subtilis* *B. subtilis*	SW P21191 SW P33449 EM L32599	1.00E-73 1.60E-89 4.60E-92

E. faecalis sequence identifier*	Homologue	Organism	Accession #	E value[a]
Mf1 (SEQ ID NO: 52) (Contig 6336)	Ptr	*S. pristinaespiralis*	X84072	4.00E-59
	LfrA	*M. smegmatis*	U40487	5.20E-51
	Mmr	*B. subtilis*	Q00538	1.40E-48
	QacA	*S. aureus*	GB X56628	1.80E-48
	SmvA	*S. typhimurium*	P37594	9.00E-45
	EmrB	*E. coli*	BAA16553	1.40E-39
	Bmr3	*B. subtilis*	Z99105	2.20E-34
Mf2 (SEQ ID NO: 54) (Contig 6226)	Bmr3	*B. subtilis*	Z99105	7.80E-62
	YfiU	*B. subtilis*	BAA24461	1.80E-46
	EmrB	*E. coli*	BAA16553	4.60E-43
	Mmr	*B. subtilis*	Q00538	2.00E-42
	Ptr	*S. pristinaespiralis*	X84072	1.10E-39
	SmvA	*S. typhimurium*	P37594	1.40E-34
	LfrA	*M. smegmatis*	U40487	1.40E-32
	QacA	*S. aureus*	X56628	4.60E-30
Mf3 (SEQ ID NO: 56) (Contig 6443)	pmrA	*S. pneumoniae*	AJ007367	3.30E-85
	NorA	*S. aureus*	P21191	3.80E-27
	Blt	*B. subtilis*	L32599	2.60E-22
	Bmr	*B. subtilis*	P33449	3.10E-22
Mf4 (SEQ ID NO: 58) (Contig 6236)	YcnB	*B. subtilis*	BAA09056	2.30E-129
	LmrB	*B. subtilis*	BAA22228	4.10E-95
	EmrB	*E. coli*	BAA16553	8.80E-49
	Ptr	*S. pristinaespiralis*	X84072	1.90E-39
	Mmr	*B. subtilis*	Q00538	6.60E-37
	Bmr3	*B. subtilis*	Z99105	4.90E-32
	QacA	*S. aureus*	X56628	4.80E-30
Mf5 (SEQ ID NO: 60) (Contig 6480)	EmrB	*E. coli*	BAA16553	1.40E-46
	Mmr	*B. subtilis*	Q00538	2.70E-43
	Bmr3	*B. subtilis*	Z99105	7.90E-39
	QacA	*S. aureus*	X56628	5.10E-37
	LfrA	*M. smegmatis*	U40487	3.90E-31
	Bcr	*E. coli*	PR JN0659	6.20E-15
Mf6 (SEQ ID NO: 62) (Contig 6248)	LmrB	*B. subtilis*	BAA22228	1.90E-67
	YcnB	*B. subtilis*	BAA09056	4.90E-54
	Mmr	*B. subtilis*	Q00538	8.30E-29
	QacA	*S. aureus*	X56628	1.80E-27
	Bmr3	*B. subtilis*	Z99105	3.60E-26
Mf7 (SEQ ID NO: 64) (Contig 6473)	YdgK	*B. subtilis*	BAA09056	4.70E-28
	EmrD	*E. coli*	P31442	4.20E-27

E. faecalis sequence identifier*	Homologue	Organism	Accession #	E value[a]
Mate 1 (SEQ ID NO: 66) (Contig 6563)	YpnP	*B. subtilis*	P54181	1.10E-13
	YucE	*B. subtilis*	CAB12921	3.90E-10
	YdhE	*H. influenza*	P45272	9.90E-08
	YisQ	*B. subtilis*	CAA70645	6.00E-07
Smr1 (SEQ ID NO: 68) (Contig 6338)	EbrA	*B. subtilis*	CAB13614	2.60E-24
	EbrB	*B. subtilis*	CAB13613	3.70E-20

[a]The E-value relates to the Smallest Sum probability of the number of hits one can expect to see by chance when searching a database of a certain size. The analysis was performed according to Gish and States (Nature Genet. 3:266-272, 1993) and Worley et al. (Genome. Res. 5: 173-184, 1995).
*denotes contig number in TIGR database (http://www.tigr.org)

Example 2: Development of Tools to Generate Gene-Knockout Mutants

The function of newly discovered genes can be studied by gene-specific mutagenesis, commonly referred to as a gene-knockout, followed by phenotypic characterization. Gene-knockout mutants in microorganisms are usually generated by the following steps: (a) construction of a vector that contains a portion of the target gene to be inactivated, (b) introduction into the host organism of the knockout vector, (c) recombination of the knockout vector into the target gene via homologous recombination such that integration disables gene function, and (d) selection of the knockout mutant using a dominant drug selection marker (e.g., kanamycin). The aforementioned strategy has been used widely in bacteria and fungi and is well developed for use in the model Gram-negative bacterium, *E. coli*, and Gram-positive bacterium, *B. subtilis*, respectively (Link et al., J. Bacteriol. 179:6228-6237, 1997; Vagner et al., Microbiology 144:3097-3104, 1999).

Prior to the present invention, knockout mutations in *E. faecalis* were generated by conjugation and transposition (see, for example, Teng et al. Plasmid 39:182-186, 1998) or, alternatively, by insertional single gene knockouts (Qin et al., Antimicrob. Agents Chemother. 42:2883-2888, 1998). Using these methods, predominantly insertional mutations were introduced. In

this case, there would be no way for removing the antibiotic selection marker without reconstituting the target gene of interest. Thus, the disruption methods cannot be adapted for the generation of multigenic knockouts.

In order to functionally identify the genes responsible for MDR efflux in *E. faecalis*, we developed a method for disrupting genes by homologous recombination. To facilitate the performing of this method, we developed gene-disruption plasmid vectors and transformation and selection methods suitable for use in *E. faecalis*. The single-gene disruption strategy is illustrated in Fig. 1. Briefly, a fragment of the target gene is generated from the *E. faecalis* genomic DNA by PCR using primers based on the identified gene sequence. Generally, the optimal size of the partial-gene fragment is approximately 800 bp. Although each primer pair and PCR reaction required optimization to generate with fidelity the gene fragment, we successfully cloned a very significant proportion of the genes listed in Table 2, including Abc1, Abc2, Abc3, Abc4, Abc5, Abc6, Abc7, Abc8, Abc8, Abc10, Abc11, Abc13, Abc14, Abc15, Abc16, Abc17, Abc20, Abc22, Abc 23, Mf1, Mf2, Mf3, Mf4, Mf5, Mf6, Mf7, and Mate1. PCR-generated gene fragments are then cloned into the *BamHI* site of the pBS/Kan plasmid vector and transformed into *E. coli* for selection on kanamycin. Each vector construction is validated by restriction digest mapping using standard methods. The knockout vectors are each individually introduced into *E. faecalis* by electroporation into the wild-type strain OGRF1 (see below for protocol). Since the pBS/Kan vector is designed for replication in *E. coli* and does not contain a Gram-positive origin of replication, it fails to autonomously replicate in Gram-positive bacteria (Teng et al., Plasmid 39:182-186, 1998). Hence, *E. faecalis* transformants selected on kanamycin are presumed to have the vector integrated into the chromosome, as illustrated in Fig. 1.

We developed methods to introduce the gene-disruption vectors into *E. faecalis* to generate gene-disruption mutants. Briefly, electrocompetent cells were prepared from a 1 liter culture of strain OGRF1 grown in Brain Heart Infusion (BHI) medium as follows. Cells were grown until late log phase, recovered by centrifugation, and the pellet washed in sterile 10% glycerol. After four washing steps, the cells were resuspended in 2 ml sterile 10% glycerol in water and stored at -70°C. Electroporation was performed with 3 μg DNA mixed with 50 μl cells in a 0.1 cm cuvette. An Invitrogen Electroporator was used with settings at 50 μF, 200 Ohm and 1.25 KV/cm. One milliliter Todd Hewitt (TH) broth was added immediately after pulsing and the cells were incubated at 37°C with agitation at 225 rpm. Transformants were selected on BHI agar containing kanamycin at 2 mg/ml and observed after 24 hours of growth. Transformation efficiency was consistently 1-6 transformants per 3 μg of DNA.

The genomic sequence in the TIGR database is derived from strain V583. Since strain V583 is a vancomycin-resistant clinical isolate, we chose for safety reasons to use the vancomycin-sensitive strain, OGRF1, for all our laboratory manipulations. There are small sequence differences between the strains, and PCR analysis indicates that OGRF1 lacks the gene products Abc9, Abc12 and Smr1.

The putative MDR-knockout mutants were examined individually utilizing a PCR-based strategy to demonstrate that integration of the knockout plasmid had occurred at the targeted site via homologous recombination. The results of this process verified that the integration events were targeted at the appropriate site in a very high proportion of the transformants examined. We generated a large panel of gene knockout mutants using the techniques described herein.

Example 3: Methods for Determining Growth Inhibition

One can readily test bacteria for growth inhibition by test substances (e.g., natural product extracts) using visual observation (MIC determination) or turbidity (absorbance at 650 nm; IC_{50} determination). Growth inhibition is calculated as the percentage decrease in absorbency, compared to the positive control cells (e.g., the same bacterial strain in the absence of any test substance). We have developed a 96-well microplate-based assay that is suitable for drug-susceptibility testing of *E. faecalis*. In this assay, *E. faecalis* is inoculated into BHI broth at $OD_{650} = 0.08$ to 0.1. This stock cell suspension is diluted 1:300 into BHI medium, and 98 μl of the cell suspension is added to 2 μl of antibacterial agent in a microtiter plate and grown overnight at 37°C. When the disruption mutants are being tested, the appropriate antibiotic (2 mg/ml kanamycin) is added into the BHI medium. Turbidity readings are measured at OD 650nm and used to calculate the IC_{50} values. MICs are also recorded at this time.

An example of the foregoing method now follows, in which single-gene disruption mutants were contacted with one of more than 40 different antimicrobial agents from disparate chemical classes. A large proportion of the single-gene disruption mutants did not exhibit sensitivity in comparison with the wild-type strain, OG1RF. In *S. cerevisiae*, it is known that several MDR pumps have overlapping sets of substrate specificities so that single-MDR-gene disruption strains may exhibit marginal drug sensitivity. In contrast, multiple-MDR-gene knockout strains exhibit markedly increased hypersensitivity as compared to the respective single-MDR-gene disruption (Mahe et al., J. Biol. Chem. 271:25167-25172, 1996; Fleckenstein et al., Yeast 15:133-137, 1999). We now demonstrate that multiple-MDR-gene disruption strains of *E. faecalis* also exhibit increased hypersensitivity as compared to single MDR-gene knockout mutants (see Example 4). Moreover, using the

single-gene disruption mutants, several MDR-efflux pumps were identified in *E. faecalis* along with their spectrum of drug substrates (Table 3). These MDR efflux pumps required functional analysis for unequivocal identification since, in most cases, sequence comparison analysis alone was inadequate. For example, *Abc7*, *Abc16*, and *Abc23* are not particularly distinct from the numerous other ABC genes (e.g., *B. subtilis NatA*; Cheng et al., Mol. Microbiol. 23:1107-1120, 1997), and the sequence similarity alone gives no indication that these transporters are MDR efflux pumps (Table 2).

The single-gene MDR-disruption strains were tested against known antibacterial agents at multiple drug concentrations to define the MIC (Table 3). A subset of the single-gene MDR-disruption strains that exhibited drug sensitivity was selected for more detailed analysis and the dose-response curve and IC_{50} were determined for each. A comparison of the growth of gene-disruption strains *ΔNorA*, *ΔAbc7*, *ΔAbc16* and *ΔAbc23* versus the wild-type strain in the presence and absence of antibacterial agents is shown in Fig. 2.

Table 3

Gene knockout	Drug	MIC (μg/ml)		WT/Δgene ratio
		wild-type strain	Δgene	
ΔAbc7	Doxorubicin	>100	1.56	64
	Daunorubicin	25	3.12	8
	Ethidium Bromide	25	6.25	4
	Ofloxacin	12.5	3.12	4
	Erythromycin	1.56	0.78	2
	Gentamicin	50	25	2
	Mithramycin	0.2	.01	2
	Norfloxacin	6.25	3.12	2
	Novobiocin	12.5	6.25	2
	Sanguinarine	12.5	6.25	2
ΔAbc16	Erythromycin	1.56	0.1	16
	Daunorubicin	25	12.5	2
	Sanguinarine	12.5	6.25	2
ΔAbc23	Clindamycin	25	0.05	500
	Lincomycin	50	0.1	500
	Virginiamycin	100	1.56	64
	Synercid	25	0.39	64
ΔNorA	Acriflavin	12.5	6.25	2
	Ciprofloxacin	1.56	0.78	2
	Clindamycin	50	25	2
	Erythromycin	1.56	0.78	2
	Norfloxacin	6.25	3.12	2
	Novobiocin	12.5	6.25	2
Bmr	Clindamycin	50	25	2
	Erythromycin	1.56	0.78	2

Example 4: Generation of Multiple Gene-Knockouts in a Single Strain

The single-gene disruption strategy described in Example 2 precludes the iterative use of the pBS/Kan vector to generate multiple insertion mutations in a single strain because of the presence of the kanamycin selectable marker following the first knockout. Furthermore, the re-use of the same plasmid backbone would likely result in homologous recombination in any subsequently introduced plasmid with shared sequence. To circumvent the foregoing problems, we developed an alternative strategy to develop methods for the generation of multiple site-specific disruption mutations in a single strain. We developed molecular tools to enable the facile generation of

site-directed multiple-knockout mutations utilizing an iterative process of (i) plasmid insertion and selection followed by (ii) screening for excision and target-gene deletion. The plasmid vectors and gene-deletion scheme are illustrated in Fig. 3 and described in detail below.

Targeted in-frame deletions are constructed in three steps. In the first step, a vector is constructed that contains sequences flanking the region to be deleted. In the second step, this vector is inserted into the bacterial chromosome by homologous recombination (referred to as a crossover) with one of these flanking regions (sequence A or sequence B in Fig. 3). Bacteria in which the first crossover has occurred are selected for their resistance to kanamycin. In the third step, a second crossover occurs with the other flanking region, resulting in loss of the inserted vector (including the gene conferring resistance to kanamycin) and of the target gene in the chromosome (Fig. 3). These three steps are described in more detail below.

Step 1. Construction of a Vector with Flanking Homologies by Crossover PCR

Primers are designed that amplify regions flanking the gene of interest. PCR products of both flanking regions are preferably at least 500 bp, and, even more preferably, at least 1 kb. PCR products should contain bases that overlap between them so that they can hybridize in the second round of PCR, plus an integral number of codons from the gene of interest to prevent polarity in the deletion. In one example, the following bases are added to the 3' primer on the N-terminal end of the gene:
5'-CCCATCCACTAAACTTAAACA -3' (SEQ ID NO: 69), while these following bases are added to the 5' primer on the C-terminal end of the gene:
5'-TCTTTAAGTTTAGTGGATGGG -3' (SEQ ID NO: 70). During the first round of PCR, two products will be made, one for each flanking region. The products are isolated on a gel to select the correct sized product.

During a second PCR step, 1 μl of each flanking region product is combined and used as a template. In this reaction, only the outside primers are used (i.e., the 5' primer from the amplification if the upstream fragment and the 3' primer from the amplification of the downstream fragment), and at a 10-fold higher concentration of each. The templates anneal to each other by their complementary bases and, thus, the outside primers will amplify the combined product (sequences A and B joined together). The product is separated on a gel, the correct-sized band is excised, and the isolated DNA is ligated into the vector.

Step 2. Insertion into the Chromosome

The completed vector, including the insert, is transformed into bacterial cells using the standard electroporation protocol with selection for kanamycin resistance. PCR and/or Southern blot may be used to verify targeted integration of this construct into the chromosome.

Step 3. Isolation of Deletion Mutants

This protocol is based on replicate plating of the strain with the inserted construct. In one example, cultures are first grown overnight in BHI at 37°C until they reach stationary phase. The overnight culture is diluted 100-fold in fresh media and grown at 37°C for an additional 90 minutes. One hundred microliters from 5- to 10-fold dilution is plated to obtain approximately 200 colonies on a TH agar plate (per liter: 30 g TH broth, 15 g agar). The colonies are transferred from the original plate onto a piece of sterile velvet cloth using a replica plating block and a metal clamp which has been pre-marked on one side for orientation. The cells are transferred from the velvet cloth, first onto a TH agar plate containing the required antibiotic marker and then to a plain TH agar plate. All three plates are marked on the

corresponding sides to align them with the marking on the metal clamp. The plates are then incubated at 37°C for 16 hours.

The marking on the TH agar plate is aligned with the marking on the TH agar selection plate and colonies growing only on the TH agar plate and not on the selection plate are identified. These cells may be plated onto fresh TH agar plates as well as TH agar selection plates for retest. Colonies that grow only on TH agar and not on the selection plates are chosen and checked for the specific deletion by PCR and Southern blot hybridization.

In the case of an MDR gene deletion mutant, the drug sensitivity phenotype is verified by the determination of drug MIC and/or IC_{50} in comparison to that of the wild-type parent strain.

The vectors and methodology for generating multiple MDR knockouts were used to generate a mutant with a deletion in the *NorA* gene. This *ΔNorA* mutant can serve as a recipient in transformation experiments using all the single-gene knockout constructs listed in Table 2. Hence, using the methods developed herein we can now rapidly generate two-gene MDR knockout mutants. We have generated such mutants (*ΔNorAΔAbc1*, *ΔNorAΔAbc2*, *ΔNorAΔAbc5*, *ΔNorAΔAbc6*, *ΔNorAΔAbc7*, *ΔNorAΔAbc8*, *ΔNorAΔAbc10*, *ΔNorAΔAbc13*, *ΔNorAΔAbc16*, *ΔNorAΔAbc17*, *ΔNorAΔAbc22*, *ΔNorAΔAbc23*, *ΔNorAΔMf2*, *ΔNorAΔMf3*, *ΔNorAΔMf6*, *ΔNorAΔf7*, and *ΔNorAΔMate1*). Significantly, the *ΔNorAΔAbc16* double mutant has the predicted phenotype since it exhibits increased dual sensitivity to norfloxacin and erythromycin (attributed to the combination of *ΔNorA* and *ΔAbc16*, respectively). The *ΔNorAΔAbc7* and *ΔNorAΔAbc23* double mutants also exhibit increased drug sensitivity.

The foregoing method can be repeated an unlimited number of times, thus allowing for the generation of multigene-knockout bacteria (e.g., *E. faecalis*). These bacteria are, in turn, useful for screening of test substances for

the identification of novel antibacterial agents, as described in Example 5. The method is readily adaptable to any bacterium in which homologous recombination can occur.

Example 5: Use of Multi-MDR Gene-knockout Bacteria for Identification of Novel Antibacterial Agents

Bacterial strains having mutations in two or more MDR efflux pumps are likely to be hypersensitive to antibacterial agents. Accordingly, these strains are useful in assays to identify novel antibacterial compounds. Methods of making multi-MDR gene disrupted bacteria and methods for measuring bacterial cell growth or metabolism are described herein. In the screening assays of the present invention, test substances are contacted with the bacterial strains for a period of time sufficient to detect any inhibition of growth or metabolism. If the test substance that inhibits growth or metabolism of a bacterial cell is a collection of compounds, the test substance can be fractionated using any standard technique to further characterize the active compound. If the test substance is a substantially pure compound, this compound can be chemically derivatized in order to attempt to find a related molecule that has greater activity or that is pumped less efficiently. The assay can be adapted for high throughput screening. Methods suitable for high throughput are described herein.

Example 6: MDR Efflux Pumps as Drug Targets

MDR efflux pumps in bacterial cells decrease the sensitivity of the cells to antibacterial agents by pumping the agents out of the cells. Thus, a compound that blocks efflux of the agent or decreases expression of the pump will make bacterial cells more sensitive to antibacterial agents. The MDR efflux pumps described herein are accordingly novel targets for these

compounds.

Test substances can be screened for their ability to decrease drug efflux. In one example, an *E. faecalis* MDR efflux pump is contacted with a test substance, and the ability of a compound in the test substance to bind to the polypeptide is then determined. Any of a number of binding assays known in the art can be performed. In another example, the ability of a test substance to decrease expression of an MDR efflux pump is determined. The method includes the steps of: (a) providing a cell expressing the MDR efflux pump; (b) contacting the cell with the test substance; and (c) measuring expression of the MDR efflux pump in the cell. Decreased polypeptide expression, relative to a cell not contacted with the test substance, indicates that the test substance decreases expression of an MDR efflux pump. Expression of the MDR efflux pump can be determined by measuring protein levels of the MDR efflux pump or by measuring levels of RNA encoding the MDR efflux pump. Finally, efflux of an antibacterial agent by an MDR efflux pump of the present invention can be determined in whole cells in the presence and absence of a test substance. In this assay, decreased efflux in the presence of the test substance identifies the test substance as containing a compound that blocks efflux through the MDR efflux pump.

In a further example, an *E. faecalis* gene encoding an MDR efflux pump is introduced into a surrogate bacterial host (e.g., *B. subtilis*, *L. lactis*) that lacks the MDR gene. The ability of a compound in the test substance to block efflux of an antibacterial agent by the MDR efflux pump from the surrogate bacterial host can be ascertained by measuring sensitivity to the antibacterial agent as measured by cell growth. Hence, the MDR efflux pump can be expressed in a heterologous bacterial expression host such that the introduced MDR efflux pump is the only *E. faecalis*-derived gene expressed in the surrogate bacterial host.

Compounds identified as blocking MDR pump efflux can be used to increase the sensitivity of an *E. faecalis* cell to an antibacterial agent. For example, the cell can be contacted with a compound that blocks efflux of the antibacterial agent from the cell by binding to an MDR efflux pump. Alternatively, the cell can be contacted with a compound that blocks efflux of the antibacterial agent from the cell by decreasing the expression of an MDR efflux pump.

Example 7: Expression of bacterial MDR efflux pumps in non-bacterial cells

The MDR efflux pumps of the present invention can be expressed in non-bacterial cells such as yeast cells or mammalian cells (e.g., human cells) in order to screen for compounds that block their efflux of antibacterial agents. For example, the *LmrA* MDR gene from *L. lactus* has been functionally expressed in GM0637 human lung fibroblast cells (van Veen et al., Nature 391:391-295, 1998). The lung fibroblast cells expressing the LmrA protein showed a 10-60 fold increased resistance to a variety of natural product drugs and synthetic chemotherapeutic drugs, which are typical substrates of human P-glycoprotein. The functional complementation of the human MDR P-glycoprotein with an expressed bacterial MDR gene allows the use of numerous methods for the discovery and characterization of agents that inhibit bacterial MDR efflux in mammalian-cell based assays (reviewed in Sharom, J. Memb. Biol.160:161-175, 1997 and Stein, Physiol. Rev. 77:545-590, 1997, each of which is hereby incorporated by reference). These approaches involve various assay formats for the measurement of the modulation by MDR "reversal agents" or blockers of compound efflux mediated by MDR pumps. Examples include the measurement, in the presence and absence of a test substance, of: (i) cytotoxicity mediated by an MDR-pump substrate (e.g. doxorubicin); (ii) cellular uptake of a radiolabeled MDR-pump substrate (e.g.

[^{3}H]daunorubicin); and (iii) accumulation of a fluorescent MDR-pump substrate (e.g. Rhodamine 123) (Bosch et al., Leukemia 11:1131-1137, 1997).

In another example, the bacterial MDR gene, *EmrE* from *E. coli*, has been expressed in the yeast *Saccharomyces cerevisiae* and shown to confer resistance to a wide variety of drugs, including acriflavin, ethidium, and methyl viologen (Yelin et al., J. Bacteriol. 181:949-956, 1999). Although the *EmrE* protein pumped drugs into the vacuolar compartment rather than into the exterior of the cell, this report further indicates the feasibility of expressing bacterial MDR genes in a non-bacterial background. As with the mammalian cell-based MDR expression system described previously, lower-eukaryote heterologous expression hosts can also be used as the basis for the discovery of MDR inhibition agents.

Example 8: Cell-free assays for the identification of compounds that block MDR efflux pumps

In addition to performing screening assays in non-bacterial cells, assays can also be performed using cell free assays. The bacterial MDR LmrP protein from *L. lactus* has been purified, reconstituted in dodecylmaltoside-destabilized preformed liposomes, and shown to mediate the transport of multiple drugs in response to an artificially-imposed pH gradient (Putman et al., Biochemistry 38:1002-1008, (1999). These studies demonstrate the applicability of using cell-free systems for the discovery of blockers of drug efflux mediated by MDR proteins. The transport assays described above (e.g., using radiolabeled [^{3}H]daunorubicin or Rhodamine 123) can be readily adapted to the cell free assays.

Test Substances

In general, test substances and compounds are identified from large libraries of both natural product and synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of the test substance is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction, fractionation, or synthetic methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their anti-pathogenic activity should be employed whenever possible.

When a test substance, as a mixture, is found to inhibit growth or metabolism of a bacterial cell, further fractionation of the positive lead extract may be necessary to isolate chemical constituents responsible for the observed

effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity having the desired activity within the test substance. Methods of fractionation and purification of test substances are known in the art. If desired, test substances shown to be useful agents for the inhibition of bacterial growth or metabolism are chemically modified according to methods known in the art.

Pharmaceutical Therapeutics and Plant Protectants

The invention provides a simple means for identifying compounds (including peptides, small molecule inhibitors, and mimetics) capable of inhibiting bacterial growth or metabolism. Accordingly, antibacterial agents discovered to have medicinal or agricultural value using the methods described herein are useful as either drugs, plant protectants, or as information for structural modification of existing anti-bacterial compounds, e.g., by rational drug design. Such methods are useful for screening compounds having an effect on a variety of bacteria.

For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Suitable routes of administration include, for example, orally, by inhalation, or by subcutaneous, intravenous, interperitoneal, intramuscular, or intradermal injections which provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of an anti-bacterial agent in a physiologically-acceptable carrier. The amount of the anti-bacterial agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the type of disease and extensiveness of the disease. Generally, amounts will be in the range of those used for other

agents used in the treatment of other bacterial diseases, although in certain instances lower amounts will be needed because of the increased specificity of the compound. A compound is administered at a dosage that inhibits bacterial growth or metabolism. For example, for systemic administration a compound is administered typically in the range of 0.1 μg to 10 g/kg body weight.

In addition, the antibacterial agent may be added to materials used to make catheters, including but not limited to intravenous, urinary, intraperitoneal, ventricular, spinal and surgical drainage catheters, in order to prevent colonization and systemic seeding by potential pathogens. Similarly, the antibacterial agent may be added to the materials that constitute various surgical prostheses and to dentures to prevent colonization by pathogens and thereby prevent more serious invasive infection or systemic seeding by pathogens.

For agricultural uses, the compositions or agents identified using the methods disclosed herein may be used as chemicals applied as sprays or dusts on the foliage of plants, or in irrigation systems. Typically, such agents are to be administered on the surface of the plant in advance of the pathogen in order to prevent infection. Seeds, bulbs, roots, tubers, and corms are also treated to prevent pathogenic attack after planting by controlling pathogens carried on them or existing in the soil at the planting site. Soil to be planted with vegetables, ornamentals, shrubs, or trees can also be treated with chemical fumigants for control of a variety of bacterial pathogens. Treatment can be done several days or weeks before planting. The chemicals can be applied by either a mechanized route, e.g., a tractor or with hand applications. In addition, chemicals identified using the methods of the assay can be used as disinfectants.

Other Embodiments

The present invention has been described in terms of particular embodiments found or proposed by the present inventors to include exemplary modes for the practice of the invention. It will be appreciated by those of skill in the art that, in light of the present disclosure, numerous modifications and changes can be made in the particular embodiments exemplified without departing from the intended scope of the invention. All such modifications are intended to be included within the scope of the invention.

What we claim is:

1. A method of determining whether a candidate nucleotide sequence encodes an MDR efflux pump, said method comprising the steps of:

(a) searching a database of nucleotide sequences for sequences having high identity to a sequence encoding a known MDR efflux pump to generate a first set of one or more candidate sequences comprising sequences that have high identity to the sequence encoding the known MDR efflux pump;

(b) from the first set of candidate sequences, selecting one or more sequences that include a sequence encoding one or more potential transmembrane domains to generate a second set of candidate sequences;

(c) in a bacterial cell, mutating a gene corresponding to one of the second set of candidate sequences; and

(d) determining whether the bacterial cell exhibits increased sensitivity to an antibacterial agent, wherein the candidate sequence encodes an MDR efflux pump if the bacterial cell exhibits increased sensitivity to the antibacterial agent.

2. A method of determining whether a polypeptide sequence functions as an MDR efflux pump, said method comprising the steps of:

(a) searching a database of polypeptide sequences for sequences having high identity to a polypeptide sequence known to function as an MDR efflux pump to generate a first set of one or more candidate polypeptides;

(b) from the first set of candidates, selecting one or more that have potential transmembrane domains to generate a second set of candidate polypeptides;

(c) in a bacterial cell, mutating a gene encoding one of the second set of candidate polypeptides; and

(d) determining whether the bacterial cell exhibits decreased sensitivity to antibacterial agents, wherein the candidate polypeptide functions as an MDR efflux pump if the cell exhibits decreased sensitivity to an antibacterial agent.

3. A method for deleting a desired region of DNA in a bacterial cell, said method comprising the steps of:

(a) transforming bacterial cells with a vector comprising (i) a first region of at least 30 nucleotides substantially identical to a first region of chromosomal DNA in the bacterial cells; (ii) a second region of at least 30 nucleotides substantially identical to a second region of chromosomal DNA in the bacterial cells; and (iii) a third region encoding a polypeptide that provides resistance to a selection agent, wherein the first and second regions of chromosomal DNA are on opposite sides of the region of DNA to be deleted;

(b) selecting for bacterial cells in which the vector of step (a) has integrated by a first crossover event into the chromosomal DNA by adding the selection agent to the culture medium; and

(c) culturing the cells selected for in step (b) in the absence of a selection agent to allow for a second crossover event to occur in at least a subset of the cells, wherein the second crossover event results in the loss of (i) the desired region of DNA and (ii) the region encoding the polypeptide that provides resistance to the selection agent in that subset of cells.

4. The method of claim 3, wherein steps (a)-(c) are repeated one or more times to delete additional regions of DNA in the same bacterial cell from which the first region was deleted.

5. A method for deleting two regions of DNA in a bacterial cell, said method comprising the steps of:

(a) transforming bacterial cells with a vector comprising (i) a first region of at least 30 nucleotides substantially identical to a first region of chromosomal DNA in the bacterial cells; (ii) a second region of at least 30 nucleotides substantially identical to a second region of chromosomal DNA in the bacterial cells; and (iii) a third region encoding a polypeptide that provides resistance to a selection agent, wherein the first and second regions of chromosomal DNA are on opposite sides of a first region of DNA to be deleted;

(b) selecting for bacterial cells in which the vector of step (a) has integrated by a first crossover event into the chromosomal DNA by adding the selection agent to the culture medium;

(c) culturing the cells selected for in step (b) in the absence of a selection agent to allow for a second crossover event to occur in at least a subset of the cells, wherein the second crossover event results in the loss of (i) the first region of DNA and (ii) the region encoding the polypeptide that provides resistance to the selection agent in that subset of cells;

(d) transforming bacterial cells resulting from step (c) with a vector comprising (i) a first region of at least 30 nucleotides substantially identical to a first region of chromosomal DNA in the bacterial cells; (ii) a second region of at least 30 nucleotides substantially identical to a second region of chromosomal DNA in the bacterial cells; and (iii) a third region encoding a polypeptide that provides resistance to a selection agent, wherein the first and second regions of chromosomal DNA are on opposite sides of a second region of DNA to be deleted;

(e) selecting for bacterial cells in which the vector of step (d) has integrated by a first crossover event into the chromosomal DNA by adding the

selection agent to the culture medium;

(f) culturing the cells selected for in step (e) in the absence of a selection agent to allow for a second crossover event to occur in at least a subset of the cells, wherein the second crossover event results in the loss of (i) the second region of DNA and (ii) the region encoding the polypeptide that provides resistance to the selection agent in that subset of cells.

6. The method of claim 5, wherein said bacterial cell is selected from the group consisting of *Staphylococcus aureus*, *Streptococcus pyogenes*, *Bacillus anthracis*, *Clostridium tetani*, *Clostridium bolulinum*, *Vibrio cholerae*, *Helicobacter pylori*, *Salmonella typhimurium*, *Shigella dysenteriae*, *Bordetella pertussis*, *Yersinia pestis*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*, *Mycobacterium tuberculosis*, *Corynebacterium diptheriae*, *Borrelia burdorferi*, *Treponema pallidum*, *Enterococcus faecalis*, *Enterococcus faecium*, and *Streptococcus pneumoniae*.

7. A method for determining whether a test substance inhibits the growth or metabolism of cells of a strain of *Enterococcus faecalis* having a disruptive mutation in a gene encoding a protein selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68), said method comprising the steps of:

(a) contacting the cells with a test substance; and

(b) determining whether the growth or metabolism of the cells is inhibited, wherein an inhibition of growth or metabolism, relative to the growth or metabolism of cells of the strain of bacteria having the mutation but not contacted with the test substance, identifies the test substance as one having the potential use as an antibacterial agent.

8. The method of claim 7, wherein the bacterial strain has disruptive mutations in genes encoding at least two of the proteins selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

9. A substantially pure nucleic acid molecule consisting essentially of *Enterococcus faecalis Abc1* (SEQ ID NO: 1), *Abc2* (SEQ ID NO: 3), *Abc3* (SEQ ID NO: 5), *Abc4* (SEQ ID NO: 7), *Abc5* (SEQ ID NO: 9), *Abc6* (SEQ ID NO: 11), *Abc7* (SEQ ID NO: 13), *Abc8* (SEQ ID NO: 15), *Abc9* (SEQ ID NO: 17), *Abc10* (SEQ ID NO: 19), *Abc11* (SEQ ID NO: 21), *Abc12* (SEQ ID NO: 23), *Abc13* (SEQ ID NO: 25), *Abc14* (SEQ ID NO: 27), *Abc15* (SEQ ID NO: 29), *Abc16* (SEQ ID NO: 31), *Abc17* (SEQ ID NO: 33), *Abc18* (SEQ ID NO: 35), *Abc19* (SEQ ID NO: 37), *Abc20* (SEQ ID NO: 39), *Abc21* (SEQ ID NO: 41), *Abc22* (SEQ ID NO: 43), *Abc23* (SEQ ID NO: 45), *Bmr* (SEQ ID NO: 47), *NorA* (SEQ ID NO: 49), *Mf1* (SEQ ID NO: 51), *Mf2* (SEQ ID NO: 53), *Mf3* (SEQ ID NO: 55), *Mf4* (SEQ ID NO: 57), *Mf5* (SEQ ID NO: 59), *Mf6* (SEQ ID NO: 61), *Mf7* (SEQ ID NO: 63), *Mate1* (SEQ ID NO: 65), or *Smr1* (SEQ ID NO: 67).

10. A substantially pure polypeptide comprising a sequence selected from the group consisting of *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

11. An *Enterococcus faecalis* bacterial cell having a disruptive mutation in a gene encoding a protein selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

12. The bacterial cell of claim 11, wherein the bacterial strain has disruptive mutations in genes encoding at least two of the proteins selected from the group consisting of Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

13. A method for determining whether a test substance includes a compound that potentially blocks efflux of antibacterial agents from a cell, said method comprising the steps of:

(a) providing a polypeptide selected from the group consisting of *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68);

(b) contacting the polypeptide with the test substance; and

(c) determining whether a compound in the test substance binds to the polypeptide, wherein binding of the compound to the polypeptide identifies it as a compound that potentially blocks efflux of antibacterial agents from a cell.

14. The method of claim 13, wherein said method is performed *in vitro*.

15. The method of claim 13, wherein said method is performed *in vivo*.

16. The method of claim 13, wherein said polypeptide is expressed in a strain other than a strain of *Enterococcus faecalis*.

17. The method of claim 13, wherein the strain is a strain of *Bacillus subtilis* or *Lactococcus lactis*.

18. A method for determining whether a test substance includes a compound that reduces efflux of antibacterial agents from a cell, said method comprising the steps of:

(a) providing a cell expressing a gene encoding an MDR efflux pump selected from *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68);

(b) contacting the cell with the test substance; and

(c) measuring expression of the gene encoding the MDR efflux pump in the cell, wherein decreased expression, relative to a cell not contacted with the test substance, identifies the test substance as containing a compound that reduces efflux of antibacterial agents from a cell.

19. The method of claim 18, wherein expression of the gene encoding the MDR efflux pump is determined by measuring protein levels of the MDR efflux pump.

20. The method of claim 18, wherein expression of gene encoding the MDR efflux pump is determined by measuring levels of RNA encoding the MDR efflux pump.

21. The method of claim 18, wherein said MDR efflux pump is expressed in a strain other than a strain of *E. faecalis*.

22. The method of claim 21, wherein the strain is a strain of *B. subtilis* or *L. lactis*.

23. A method for increasing the sensitivity of an *E. faecalis* cell to an antibacterial agent, said method comprising the step of contacting the cell with a compound that blocks efflux of the antibacterial agent from the cell by binding to an MDR efflux pump selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

24. A method for increasing the sensitivity of an *E. faecalis* cell to an antibacterial agent, said method comprising the step of contacting the cell with a compound that blocks efflux of the antibacterial agent from the cell by decreasing the expression of a gene encoding an MDR efflux pump selected from the group consisting of *E. faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO: 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68).

25. A method for determining whether a test substance functions as an inhibitor of an MDR efflux pump, said method comprising the steps of:

(a) expressing a nucleic acid molecule encoding an MDR efflux pump derived from a first bacterial strain in a second bacterial strain to provide increased resistance to an antibacterial agent relative to the second bacterial strain not expressing the MDR efflux pump;

(b) contacting the second bacterial strain with an amount of the antibacterial agent that inhibits growth of a control strain but does not substantially inhibit growth of the second bacterial strain;

(c) contacting the second bacterial strain with a test substance; and

(d) measuring growth of second bacterial strain, wherein decreased growth of the bacterial strain identifies the test substance as one that includes an inhibitor of an MDR efflux pump.

26. The method of claim 25, wherein the second bacterial strain is a strain of *Bacillus subtilis* or *Lactococcus lactis*.

27. A method for determining whether a test substance includes a compound that reduces efflux of antibacterial agents from a cell, said method comprising the steps of:

(a) providing a non-bacterial cell expressing a nucleic acid molecule encoding an MDR efflux pump selected from *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID NO 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ ID NO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68);

(b) contacting the cell with the test substance;

(c) contacting the cell with a compound that is capable of (i) entering the cell and (ii) being transported by the MDR efflux pump; and

(d) measuring efflux of the compound from the cell, wherein decreased efflux, relative to a cell not contacted with the test substance, identifies the test substance as containing a compound that reduces efflux of antibacterial agents from a cell.

28. The method of claim 27, wherein the cell is a eukaryotic cell.

29. The method of claim 28, wherein the cell is a mammalian cell.

30. The method of claim 28, wherein the cell is an insect cell.

31. The method of claim 28, wherein the cell is a yeast cell.

32. The method of claim 31, wherein the yeast is *Saccharomyces cerevisiae*.

33. A method for determining whether a test substance includes a compound that reduces efflux of antibacterial agents from a cell, said method comprising the steps of:

(a) providing a cell free system consisting of:

(i) a lipid membrane into which is inserted an MDR efflux pump selected from *Enterococcus faecalis* Abc1 (SEQ ID NO: 2), Abc2 (SEQ ID NO: 4), Abc3 (SEQ ID NO: 6), Abc4 (SEQ ID NO: 8), Abc5 (SEQ ID 10), Abc6 (SEQ ID NO: 12), Abc7 (SEQ ID NO: 14), Abc8 (SEQ ID NO: 16), Abc9 (SEQ ID NO: 18), Abc10 (SEQ ID NO: 20), Abc11 (SEQ ID NO: 22), Abc12 (SEQ ID NO: 24), Abc13 (SEQ ID NO: 26), Abc14 (SEQ ID NO: 28), Abc15 (SEQ ID NO: 30), Abc16 (SEQ IDNO: 32), Abc17 (SEQ ID NO: 34), Abc18 (SEQ ID NO: 36), Abc19 (SEQ ID NO: 38), Abc20 (SEQ ID NO: 40), Abc21 (SEQ ID NO: 42), Abc22 (SEQ ID NO: 44), Abc23 (SEQ ID NO: 46), Bmr (SEQ ID NO: 48), NorA (SEQ ID NO: 50), Mf1 (SEQ ID NO: 52), Mf2 (SEQ ID NO: 54), Mf3 (SEQ ID NO: 56), Mf4 (SEQ ID NO: 58), Mf5 (SEQ ID NO: 60), Mf6 (SEQ ID NO: 62), Mf7 (SEQ ID NO: 64), Mate1 (SEQ ID NO: 66), and Smr1 (SEQ ID NO: 68);

(ii) the test substance;

(iii) a compound that is capable of being transported by the MDR efflux pump; and

(b) measuring transport of the compound across the lipid membrane, wherein decreased transport, relative to transport in a cell-free system not comprising the test substance, identifies the test substance as containing a compound that reduces efflux of antibacterial agents from a cell.

34. The method of claim 33, wherein said lipid membrane is in the form of a liposome.

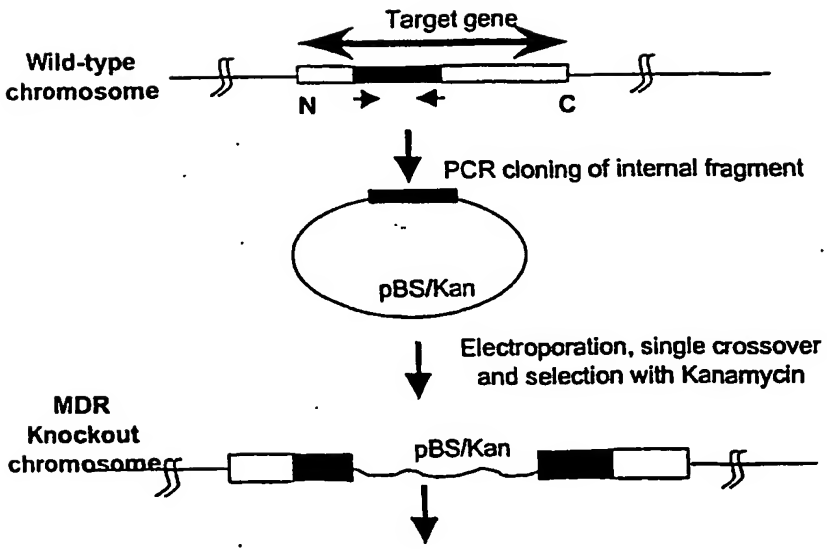

Target gene
Wild-type chromosome
N
C
PCR cloning of internal fragment
pBS/Kan
Electroporation, single crossover and selection with Kanamycin
MDR Knockout chromosome
pBS/Kan
No functional gene product due to insertional mutagenesis

Fig. 1

Fig. 2A

E. faecalis Δ*NorA* MDR pump knock-out :
3-fold More Sensitive to Norfloxacin

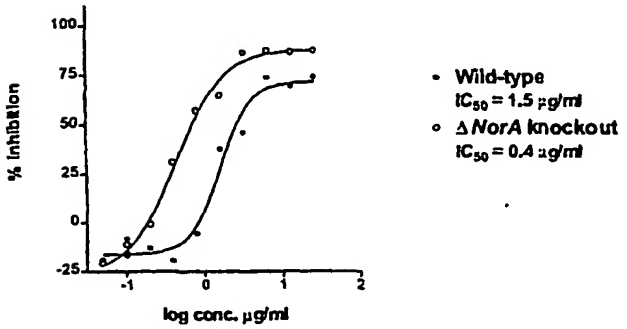


Fig. 2B

E. faecalis Δ*Abc7* MDR pump knockout :
9-fold More Sensitive to Daunorubicin

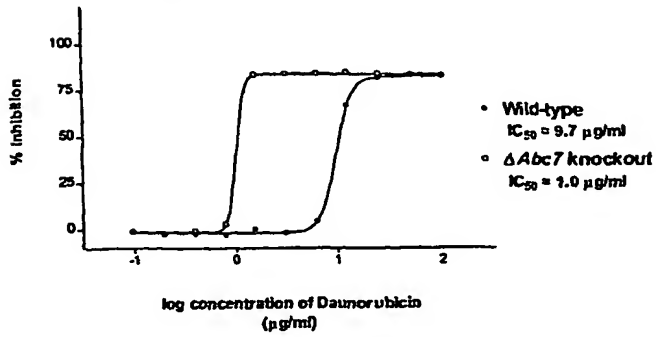


Fig. 2C

E. faecalis Δ*Abc7* MDR pump knockout :
58-fold More Sensitive to Doxorubicin

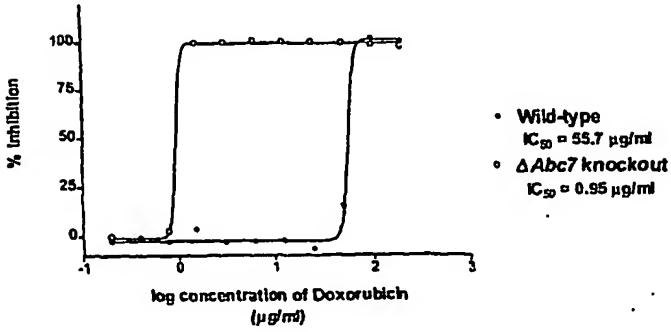


Fig. 2D

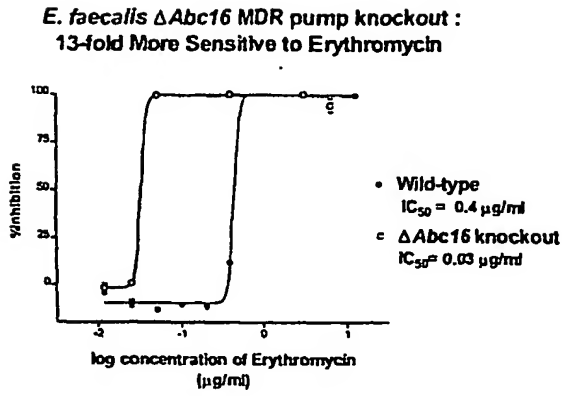


Fig. 2E

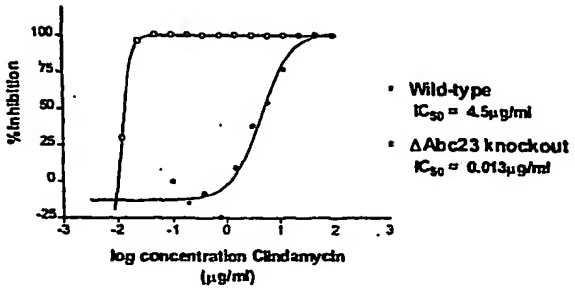
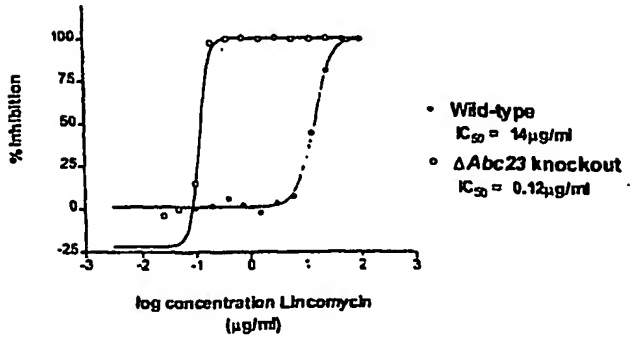
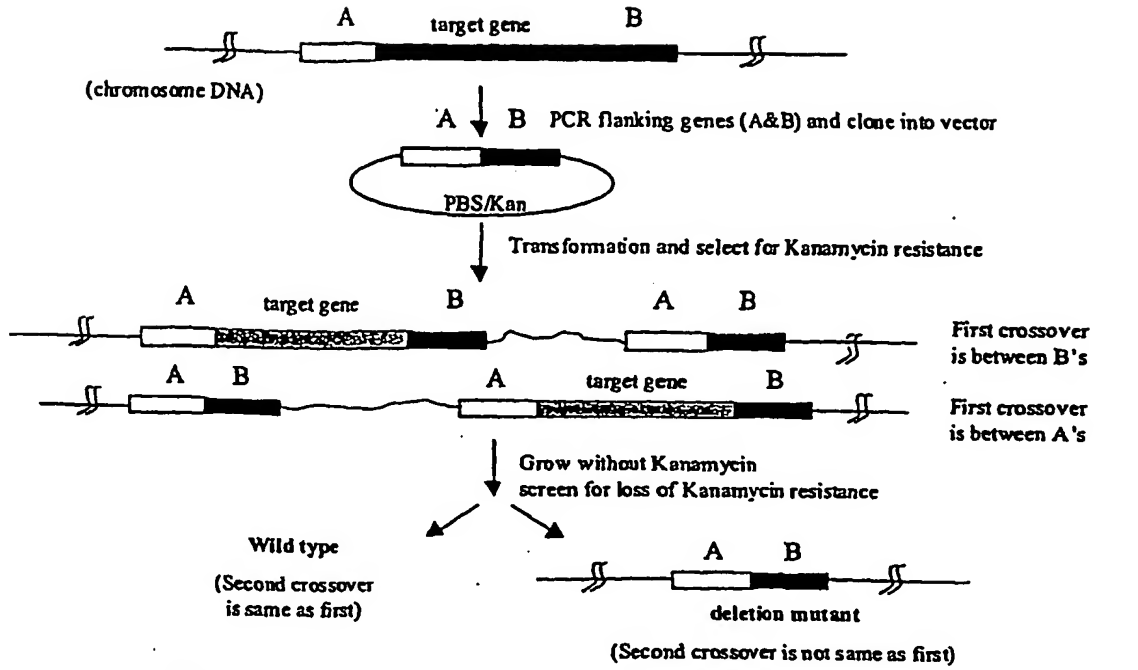


Fig. 2F





Detail of the second cross-over

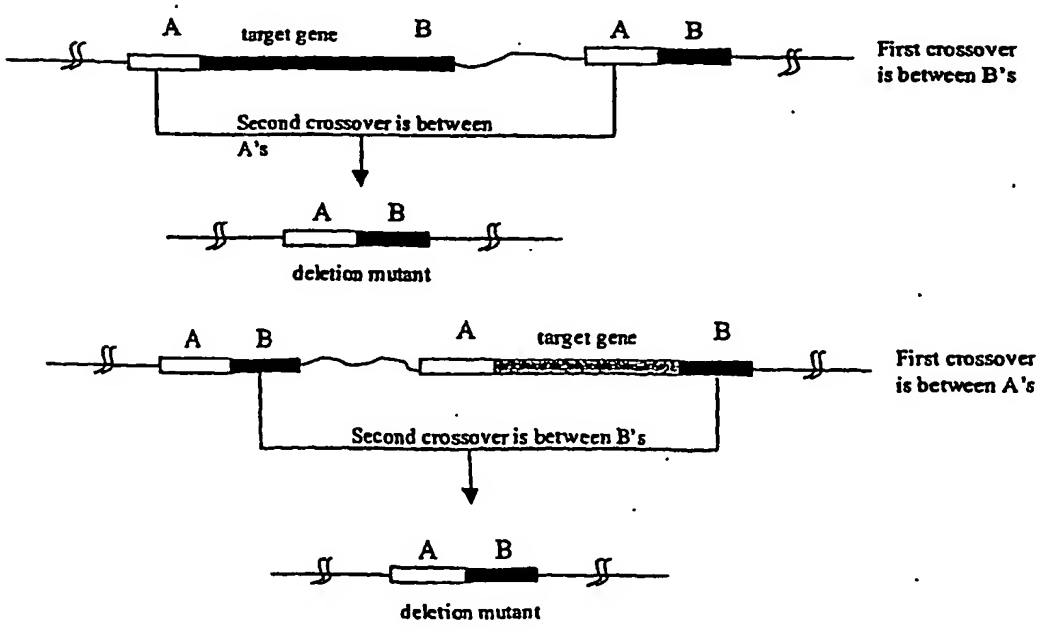


Fig. 3

```
5'  ATG TTT GGT TTA TTG AAA TAC GCA AAA AAT TAT CGA AAA CAA ATT ATC CTC GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   F   G   L   L   K   Y   A   K   N   Y   R   K   Q   I   I   L   G

    CCA GTC TTC AAA TTT TTA GAA GCT TGT TTC GAG TTA GTC TTG CCC TTA TTC ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   V   F   K   F   L   E   A   C   F   E   L   V   L   P   L   F   M

    GCT CAC TTA GTC GAT GTC GGT ATT CGT CAA AAC GAC CGA CAG ACT GTC ATA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   H   L   V   D   V   G   I   R   Q   N   D   R   Q   T   V   I   E

    ATG GCC CTC TGG ATG CTT GTT ATG TCG TTA GTT GGT TTA TTT TTT GTC ATG ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   A   L   W   M   L   V   M   S   L   V   G   L   F   F   V   M   I

    TGT CAA TAC TAT GCT TCA GTC GCT TCG CAA GGT TTT GGG ACG GAG CTA AGA AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     C   Q   Y   Y   A   S   V   A   S   Q   G   F   G   T   E   L   R   N

    CAA TTA ATG AAG AAA ATT AAT CAG CTT TCA CAC AAA GAA TTG AAT AGT TTT GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   L   M   K   K   I   N   Q   L   S   H   K   E   L   N   S   F   G

    ACA GAT ACC CTC ATC ACC CGG ATC ACA AAC GAT ATC AAC CAG CTT CAA TTA GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   D   T   L   I   T   R   I   T   N   D   I   N   Q   L   Q   L   A

    TTA GCG ATG TTT ATT CGG TTG GTC ATT CGG ACA CCT TTT TTA AGT ATC GGT TCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   A   M   F   I   R   L   V   I   R   T   P   F   L   S   I   G   S

    GTG GTG ATG GCT TTT TAC ATT GAC GTG CAG ATG GGC TTT CTT TTC CTA TTA CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   V   M   A   F   Y   I   D   V   Q   M   G   F   L   F   L   L   L

    TTA CCA ATT TTT AGC CTT ATT CTC TTT ATT ATC ATT AAA GTG ACT GTG CCT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   P   I   F   S   L   I   L   F   I   I   I   K   V   T   V   P   L

    TAT CAA AAA GTC CAA GAA TAT TTG GAT CGG TTA AAC CGT CAA ATC AGT CAA AAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Y   Q   K   V   Q   E   Y   L   D   R   L   N   R   Q   I   S   Q   N

    TTA AGC GGT GTC CGT GTG ATC CGT GCG TTT GCT AGA AAG GAA ACA GAG CAA CGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   S   G   V   R   V   I   R   A   F   A   R   K   E   T   E   Q   R

    CAT GTT GAT AAA GCT TCA GAT GAT TTA GGC GAT ATT TAC ATT CGT GTA TCG AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     H   V   D   K   A   S   D   D   L   G   D   I   Y   I   R   V   S   N

    GTC TCT GCT TTA TTA ACG CCT TTA ACC ACT TTG ATT ATG AAT GTT GGA ATT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   S   A   L   L   T   P   L   T   T   L   I   M   N   V   G   I   L

    TTT TTA CTT TAT TTT AGT GGC TTA AAA GTT TCT TTT GGT TCC TTA CAA CAA GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   L   L   Y   F   S   G   L   K   V   S   F   G   S   L   Q   Q   G

    GAA GTT TTA GCA TTG ATC AAT TAT ATG AAT CAA ATG ATG CTC GCT TTA ATT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   V   L   A   L   I   N   Y   M   N   Q   M   M   L   A   L   I   V
```

Fig. 4A

```
GCT TCT AAT CTC GTG GTA ATT TTT ACG CGA GCT GCC GCT TCC GCA AAC CGT GTC
 A   S   N   L   V   V   I   F   T   R   A   A   A   S   A   N   R   V

AAT GAG GTT TTA ACT GTA GAA AGC CAG TTA ACA GAT ACA CCA GAA TCA GCA AAA
 N   E   V   L   T   V   E   S   Q   L   T   D   T   P   E   S   A   K

ACA TCT CCA CAA TTT GGT GAT ATT ACC TTT GAC CAT GTA GAT TTT CGC TAT GAA
 T   S   P   Q   F   G   D   I   T   F   D   H   V   D   F   R   Y   E

CCA GAG GCC GGT TTG GCT TTG GAA AAT ATC AAT TTT ACG ATT CCT AAA GGC TCA
 P   E   A   G   L   A   L   E   N   I   N   F   T   I   P   K   G   S

ATT CTC GGC ATC ACA GGA CCT ACC GGC AGC GGG AAA AGT ACC TTA ACC CAA CTC
 I   L   G   I   T   G   P   T   G   S   G   K   S   T   L   T   Q   L

ATT CCC AGA TTT TAT GAT GTG AGT GCA GGA AAC CTT TTT ATT AAT GGT GTA AAT
 I   P   R   F   Y   D   V   S   A   G   N   L   F   I   N   G   V   N

GTG CGC GAT TGG CCG CTC TTT ACT TTA CGC CAA CAA GTT GCA AGT GTT CCA CAA
 V   R   D   W   P   L   F   T   L   R   Q   Q   V   A   S   V   P   Q

ACT GCC GTC TTA TTT ACA GGG ACT ATT CGA GAA AAC TTA CAA TGG GGC AAA CCA
 T   A   V   L   F   T   G   T   I   R   E   N   L   Q   W   G   K   P

AAT GCA ACT GAT GAA GAC TGT TGG GAA GCA TTA GCT ATC GCC CAA TGT AAA GAA
 N   A   T   D   E   D   C   W   E   A   L   A   I   A   Q   C   K   E

TTT GTT GAA CAA TTA GAC CAA GGG TTA GAC ACG CCT GTT AAC GAA GGC GGA AAA
 F   V   E   Q   L   D   Q   G   L   D   T   P   V   N   E   G   G   K

AAC TTT TCT GGT GGA CAA AGA CAA CGG TTG ACC ATT GCC CGT GCC TTG ATT AGG
 N   F   S   G   G   Q   R   Q   R   L   T   I   A   R   A   L   I   R

AAA CCG CAT TTA CTT ATT TTA GAT GAT TCC TTG AGT GCC CTC GAC TAT CAA ACA
 K   P   H   L   L   I   L   D   D   S   L   S   A   L   D   Y   Q   T

GAT TTG AAT CTG CGT CGT GCT TTA CAA AAA GAA CGA GCA GAG ACA ACC GTT ATT
 D   L   N   L   R   R   A   L   Q   K   E   R   A   E   T   T   V   I

TTA ATT TCA CAA CGC GTG AGT TCG ATT GCA ACG GCG AAT CAA ATT TTA GTC CTA
 L   I   S   Q   R   V   S   S   I   A   T   A   N   Q   I   L   V   L

GAT AGT GGA AAA GTC GCT GGC CTT GGC ACC CAC GAA GAA TTA CTT ACT TCT TCT
 D   S   G   K   V   A   G   L   G   T   H   E   E   L   L   T   S   S

AAA GAG TAT CAA GAA ATC GTT GCG TCA CAG GAG GAG GAT ACC CAT GCA AAC TAA 3'
 K   E   Y   Q   E   I   V   A   S   Q   E   E   D   T   H   A   N   *
```

Fig. 4B

```
5'  ATG AAA CGA ATC GGT CGC TAT ATT AAA CCG TAC CGA GTG ACG TTT TAT TTA GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   K   R   I   G   R   Y   I   K   P   Y   R   V   T   F   Y   L   V

    ATT TTA TTT ACA ATA TTA ACC GTT GCC TTT AAT GCA GCG TTG CCT TAT TTG ACT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   L   F   T   I   L   T   V   A   F   N   A   A   L   P   Y   L   T

    GGA TTA CCG ACG ACA GAA ATT AGC CGT AAT ATT GCG GCC GGC GAA TCC ATT AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   L   P   T   T   E   I   S   R   N   I   A   A   G   E   S   I   N

    TTT GAT TAT GTA ATC CAA TGT TTA ATT TGG ATT TTA GTT GTG GGA ACA GGT TAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   D   Y   V   I   Q   C   L   I   W   I   L   V   V   G   T   G   Y

    TGT GTG GCA CAA TTT TTG TCA GGC TTT TTA ATG ACG AAT GTC GTT CAA CAA TCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     C   V   A   Q   F   L   S   G   F   L   M   T   N   V   V   Q   Q   S

    ATG CGC GAT TTG CGT CGC GAT ATT GAA GAA AAA ATC AAT CGT TTG CCA GTT TCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   R   D   L   R   R   D   I   E   E   K   I   N   R   L   P   V   S

    TAT TTT GAT AAG AAC CAA CAA GGA AAT ATT TTG TCA CGG GTG ACG AAC GAT GTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Y   F   D   K   N   Q   Q   G   N   I   L   S   R   V   T   N   D   V

    GAT GCT GTC AGC AAT GCG ATG CAA CAA AGT TTT ATC AAT ATT GTT TCA GCA GTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   A   V   S   N   A   M   Q   Q   S   F   I   N   I   V   S   A   V

    TTA GGT ATT GTG ATG GCG GTA GTG ATG ATG TTC TTA ATC AAT CCG CTG ATG GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   G   I   V   M   A   V   V   M   M   F   L   I   N   P   L   M   A

    ATT TTT TCA GTG ATT ATG ATT CCG TTG TCT CTG ATT ATT TCC AGA ACA ATT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   F   S   V   I   M   I   P   L   S   L   I   I   S   R   T   I   V

    AAA ATC TCC CAG AAA TAT TTC CAA GGA ATG CAA AAT TCT TTA GGA GAC TTA AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   I   S   Q   K   Y   F   Q   G   M   Q   N   S   L   G   D   L   N

    GGT TAT GTC CAA GAA AAT ATG ACT GGG TTC AGT GTC TTA AAA CTA TAT GGT CGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   Y   V   Q   E   N   M   T   G   F   S   V   L   K   L   Y   G   R

    GAA AAA GAA ACC CTT GAA GGC TTT AAA CAA GTC AAT CAT CGT TTA AAT GGT TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   K   E   T   L   E   G   F   K   Q   V   N   H   R   L   N   G   F

    GGT TTC AAA GCA TCC TTT ATC TCA GGA TTA ATG TTG CCA TTG GTT CAG ATG ACC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   F   K   A   S   F   I   S   G   L   M   L   P   L   V   Q   M   T

    GCT TAT GGG ACC TAT ATC GGG GTA GCT GTC CTT GGT AGT TAC TAT GTG GTT GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   Y   G   T   Y   I   G   V   A   V   L   G   S   Y   Y   V   V   A

    GGT GTG ATC GTA GTG GGG CAA TTA CAA GCG TTT ATT CAA TAT ATT TGG CAA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   V   I   V   V   G   Q   L   Q   A   F   I   Q   Y   I   W   Q   I
```

Fig. 5A

```
AGC CAA CCA ATG GGG AAT ATT ACG CAG TTG TCT GCA GCT TTA CAA AGC GCT TCA
 S   Q   P   M   G   N   I   T   Q   L   S   A   A   L   Q   S   A   S

GCT TCG ACC ATG CGG ATT TTT GAA ATC CTA GAT GAA CCA GAA GAA GAA CTT AAC
 A   S   T   M   R   I   F   E   I   L   D   E   P   E   E   E   L   N

GAA CAA GAT GTT CCT TTG CCA GAA CCT ATT TTA GGC TCT GTT GAA TTT GAA AAT
 E   Q   D   V   P   L   P   E   P   I   L   G   S   V   E   F   E   N

GTC AGC TTT AGT TAT GAC CCA GAA AAA CCG TTA ATT CGT AAT TTG AAC TTT AAA
 V   S   F   S   Y   D   P   E   K   P   L   I   R   N   L   N   F   K

GTT GAT GCG GGC CAA ATG GTT GCG ATT GTG GGA CCA ACT GGC GCT GGG AAA ACA
 V   D   A   G   Q   M   V   A   I   V   G   P   T   G   A   G   K   T

ACC TTA ATC AAC TTA CTG ATG CGT TTT TAT GAT GTA ACA GAA GGC GCC ATT AAA
 T   L   I   N   L   L   M   R   F   Y   D   V   T   E   G   A   I   K

ATT GAT GGC ATT GAC ACG AAA AAA ATG AAC CGT AGT GAT GTC CGA TCT GTA TTT
 I   D   G   I   D   T   K   K   M   N   R   S   D   V   R   S   V   F

GGA ATG GTA TTG CAA GAT GCT TGG TTG TAT AAA GGT ACC ATT GCA GAT AAC ATT
 G   M   V   L   Q   D   A   W   L   Y   K   G   T   I   A   D   N   I

CGT TTT GGG AAG TTA GAT GCC ACG GAT TAT GAA GTT GTC GAT GCA GCG AAA ACG
 R   F   G   K   L   D   A   T   D   Y   E   V   V   D   A   A   K   T

GCC AAT GTG GAT CAC TTC ATT CGG ACA ATG CCA GAC GGG TAT GAA ATG GAA ATC
 A   N   V   D   H   F   I   R   T   M   P   D   G   Y   E   M   E   I

AAT TCT GAG GGA GAT AAC GTT TCC CTT GGT CAA AAA CAA TTG TTG ACC ATT GCC
 N   S   E   G   D   N   V   S   L   G   Q   K   Q   L   L   T   I   A

CGA GCG GTA ATT TCT GAT CCG AAA ATT TTG ATT TTA GAT GAG GCG ACT AGT TCA
 R   A   V   I   S   D   P   K   I   L   I   L   D   E   A   T   S   S

GTC GAT ACA CGC TTG GAA GCC TTA ATT CAA AAA GCA ATG GAT CGT GTT ATG GAA
 V   D   T   R   L   E   A   L   I   Q   K   A   M   D   R   V   M   E

GGA CGA ACG AGT TTC GTT ATT GCC CAC CGC CTA TCA ACT ATT CGT GAA GCT GAT
 G   R   T   S   F   V   I   A   H   R   L   S   T   I   R   E   A   D

TTA ATT CTT GTT ATG AAA CAA GGA GAA ATC ATT GAA AAA GGT ACG CAT CAT GAG
 L   I   L   V   M   K   Q   G   E   I   I   E   K   G   T   H   H   E

TTG CTG GAA CAA GGT GGC TTC TAT GAA AAA CTA TAC AAT AGT CAA TTT GCT GAA
 L   L   E   Q   G   G   F   Y   E   K   L   Y   N   S   Q   F   A   E
```

Fig. 5B

```
GAA GGC GAC TAT GAG GAA TAA 3'
--- --- --- --- --- --- ---
 E   G   D   Y   E   E   *
```

Fig. 5C

```
5'  ATG TTT TAC ACA GCT ATT CTC TTT GCA TTT GGC TTA TCA ATT ATC AAT ATT CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   F   Y   T   A   I   L   F   A   F   G   L   S   I   I   N   I   L

    TTG CCT AGA GTC ATT CAA GTT TTT ATG GAT AAT TAT TTA ACG CCT AAA ACT GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   P   R   V   I   Q   V   F   M   D   N   Y   L   T   P   K   T   A

    ACA ATG CAA GTG ATT CTA ACT TTT GCT GCT ATT TAT TTC TTT GGT GTC ATT GTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   M   Q   V   I   L   T   F   A   A   I   Y   F   F   G   V   I   V

    AAA AGT ATT GTT TGG TTT TTC CAA TGG TAT TTG TAT TCG ATG GCT TCT TTA AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   S   I   V   W   F   F   Q   W   Y   L   Y   S   M   A   S   L   K

    ACG TAT CAA TAT ATA CGA GTG AAA TTA TTT GAA AAA TTA CAC ACT TTA GGT ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   Y   Q   Y   I   R   V   K   L   F   E   K   L   H   T   L   G   M

    CGT TAT TTT GAT CAA ACA CCA GCG GGT TCG ACG GTG TCT CGT GTT ACA AAT GAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   Y   F   D   Q   T   P   A   G   S   T   V   S   R   V   T   N   D

    ACA GAA ACA TTG TTT GAA TTC TGG TAT GTA TTC CTA ATG GTG ATT ACA GGA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   E   T   L   F   E   F   W   Y   V   F   L   M   V   I   T   G   I

    TTT GCC GTA ATC TCG TCA TTT TTT GCA ATG TTT CAA CTC AGT CCA GAA ATC TCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   A   V   I   S   S   F   F   A   M   F   Q   L   S   P   E   I   S

    TTT TAT TGT CTA ATT TTC TTG CCG ATT TTA TTA GTT GTG ATT TGG TAT TAC CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   Y   C   L   I   F   L   P   I   L   L   V   V   I   W   Y   Y   Q

    AAA TTT AGT TCA AAA TTA TAT CGC AGT ATG CGG GAA AAA TTA AGT CAA TTA AAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   F   S   S   K   L   Y   R   S   M   R   E   K   L   S   Q   L   N

    ACC AAA CTA AAT GAA TAC ATC TCT GGT ATG CAA ATC ATT CAA CAA TTT CGC CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   K   L   N   E   Y   I   S   G   M   Q   I   I   Q   Q   F   R   Q

    GAA AAG CGT TTG GAA AAA GAA TTT GAG GAA ACC AAT GAT GAT TAT TTA AAA ACA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   K   R   L   E   K   E   F   E   E   T   N   D   D   Y   L   K   T

    CGA GTA GCA ATG ATT CGT ATG AAT TCG TTA TTA TTA AGT CCC ATC ATT AAT TTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   V   A   M   I   R   M   N   S   L   L   L   S   P   I   I   N   L

    TTG TAT ACG TTG GCG ATT GCA TTG GCC TTA ACT ATG TTT GGC ATC GAT GCT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   Y   T   L   A   I   A   L   A   L   T   M   F   G   I   D   A   L

    CAT TCA CCT GTT GAA GCG GGG ATG ATT TAT GCT TTT GTA ACG TAT GTC CAA GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     H   S   P   V   E   A   G   M   I   Y   A   F   V   T   Y   V   Q   A

    TTT TTT AAT CCG ATG ACG CAA ATG ATG GAC TTC CTA AGT ATC TTT ACA GAT GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   F   N   P   M   T   Q   M   M   D   F   L   S   I   F   T   D   G
```

Fig. 6A

```
ATT GTC GCA GGT AGC CGT ATT TTA AAA ATT ATG GAT ACG GAA GAA TTA ACG CCG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 I    V    A    G    S    R    I    L    K    I    M    D    T    E    E    L    T    P

CAA CAA TCA GTA GGT GCG AAT GGG GAA ATT ATT CGC GGG AAA ATT GAA TTT CGC
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Q    Q    S    V    G    A    N    G    E    I    I    R    G    K    I    E    F    R

AAT GTT ACT TTT TCT TAT GAC GGC AAA AAC GAA GTA TTG AAA AAT ATT AGC TTT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 N    V    T    F    S    Y    D    G    K    N    E    V    L    K    N    I    S    F

GTC GCA AAT CCC GGG GAA ACC GTC GCT TTA GTC GGT CAT ACA GGA AGT GGT AAA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 V    A    N    P    G    E    T    V    A    L    V    G    H    T    G    S    G    K

AGT TCA ATT ATT AAT GTA CTC ATG CGT TTC TAT GAA TTT TAT GAA GGG CAG ATT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 S    S    I    I    N    V    L    M    R    F    Y    E    F    Y    E    G    Q    I

TTA ATT GAT GAT CGT GAC ATT CGT GAT TTT CCC ATG ACA GAA TTA CGG GAA AAG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 L    I    D    D    R    D    I    R    D    F    P    M    T    E    L    R    E    K

ATG GGC TTA GTA TTA CAG GAC GCT TTC ATG TTT TAT GGA GAT ATT GCT GGA AAT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 M    G    L    V    L    Q    D    A    F    M    F    Y    G    D    I    A    G    N

ATC CGT TTA CTG AAT CCC AAT ATT ACA GAT GAG CAA ATT AAA CAG GCT GCG GAA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 I    R    L    L    N    P    N    I    T    D    E    Q    I    K    Q    A    A    E

TTT GTT CAG GCA GAT AAA TTT ATT CAC ACG TTA CCC AAT ACC TAT CAT GCG AAA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 F    V    Q    A    D    K    F    I    H    T    L    P    N    T    Y    H    A    K

GTC ATT GAA CGA GGA GCG AGT TAT TCT AGC GGG CAA CGT CAA TTA ATT TCG TTT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 V    I    E    R    G    A    S    Y    S    S    G    Q    R    Q    L    I    S    F

GCA CGA ACG ATT GTG ACT GAT CCC AAA ATT TTA GTT TTA GAT GAA GCA ACC GCT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 A    R    T    I    V    T    D    P    K    I    L    V    L    D    E    A    T    A

AAT ATT GAT ACA GAA ACA GAA GGA TTG ATT CAA GAA GGA TTA GCC AAA ATG CGT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 N    I    D    T    E    T    E    G    L    I    Q    E    G    L    A    K    M    R

CAA GGG AGA ACG ACA ATT GCG ATT GCG CAT CGC CTT TCG ACG ATT CGT GAC GCA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Q    G    R    T    T    I    A    I    A    H    R    L    S    T    I    R    D    A

AAC TTG ATT TTA GTT TTA GAC AAA GGG CAG ATT GTA GAA CGA GGA ACG CAC GAA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 N    L    I    L    V    L    D    K    G    Q    I    V    E    R    G    T    H    E

ACG TTA CTG GCA GAA GGT GGC CTT TAT GCG GAT ATG TAT CAG TTA CAA AGT ACG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 T    L    L    A    E    G    G    L    Y    A    D    M    Y    Q    L    Q    S    T

GAA GTT TAA 3'
---  ---  ---
 E    V    *
```

Fig. 6B

```
5'  ATG AGT AGT CGA CCA AAA GGC GGA CCA ATG GGC GGT GGA CCA GGA CGC AAC ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   S   S   R   P   K   G   G   P   M   G   G   G   P   G   R   N   I

    GGC GCA AAA GGG CCA AAA CCA AAA AAT TTC TGG AAA ACA GTG AAA CGT TTA TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   A   K   G   P   K   P   K   N   F   W   K   T   V   K   R   L   F

    CGA TAC ATG TCT AAA CGA ATG CTG TCA ATT ATT GCT GTC TTA GTA TTA GCA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   Y   M   S   K   R   M   L   S   I   I   A   V   L   V   L   A   I

    GCA GCC GTC GTT TTC CAA ATT CAA ACA CCA AAA GTA TTA GGA CAG GCA ACA ACT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   A   V   V   F   Q   I   Q   T   P   K   V   L   G   Q   A   T   T

    GAA ATT TTT AAA GGC GTC ATG AAA GGC GCT GCT GAA ATG AAA CAA GGC TTA AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   I   F   K   G   V   M   K   G   A   A   E   M   K   Q   G   L   K

    ATA ACA TCT TTT CCA ATT GAT TTT GAC AAA ATT GGT CAA ATT TTA TTA ATT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   T   S   F   P   I   D   F   D   K   I   G   Q   I   L   L   I   V

    ATT GCC ATG TAC CTT ATT TCT GCT GTC TTT AAT TTC TTG CAG CAA GTA ATT ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   A   M   Y   L   I   S   A   V   F   N   F   L   Q   Q   V   I   M

    ACA CGT GTT TCA CAG CGA ACA GTT TAT GAA TTA CGT CAA GAA TTA GAA GCG AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   R   V   S   Q   R   T   V   Y   E   L   R   Q   E   L   E   A   K

    ATG AAC AAA GTT CCA ATT TCC TAT TAT GAT ACT CAT AGT AAT GGG GAC ATT ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   N   K   V   P   I   S   Y   Y   D   T   H   S   N   G   D   I   M

    TCA CGG GCA ATT AAC GAC ATG GAT AAT ATT GCG AGT ACT TTA CAG CAA AAC TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   R   A   I   N   D   M   D   N   I   A   S   T   L   Q   Q   N   L

    ACC CAA TTA ATT ACC AGT ATC GTG ACA TTT GTG GGT GTA CTT TGG ATG ATG CTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   Q   L   I   T   S   I   V   T   F   V   G   V   L   W   M   M   L

    ACC ATT AGC TGG CAG TTA ACA CTA ATT GCC TTA GCA ACA GTG CCA TTA AGT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   I   S   W   Q   L   T   L   I   A   L   A   T   V   P   L   S   L

    ATT GTT GTG ATG GTT GTG GCG CCT CGC TCG CAA AAA CAT TTT GCG GCT CAA CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   V   V   M   V   V   A   P   R   S   Q   K   H   F   A   A   Q   Q

    AAA AGC TTA GGA TTA TTG AAT AAC CAA GTC GAA GAA ACC TAC GGT GGT CAC GTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   S   L   G   L   L   N   N   Q   V   E   E   T   Y   G   G   H   V

    GTA GTA AAA AGT TTC AAC CAC GAA GAA AGT GAT CAA GAA GTA TTT GAA AAA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   V   K   S   F   N   H   E   E   S   D   Q   E   V   F   E   K   E

    AAT GAA AAA TTA TAT CAT GCT GGT CGC AAA GCA CAA TTT ATT TCA GCG ATC ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   E   K   L   Y   H   A   G   R   K   A   Q   F   I   S   A   I   I
```

Fig. 7A

```
ATG CCT TTA ATG AAC TTT ATC AAA AAT CTA GGC TAC GTG TTT GTT GCA GTC CTT
---------------------------------------------------------------------
 M   P   L   M   N   F   I   K   N   L   G   Y   V   F   V   A   V   L

GGT GGC GTA AAA GTA GCG AAT GGT ATG ATG GAT TTA GGG GAT GTC CAA GCA TTT
---------------------------------------------------------------------
 G   G   V   K   V   A   N   G   M   M   D   L   G   D   V   Q   A   F

CTT CAA TAT ACC AAT CAA TTT TCA CAA CCG ATT ACT CAA ATC GCT AAT TTA ATG
---------------------------------------------------------------------
 L   Q   Y   T   N   Q   F   S   Q   P   I   T   Q   I   A   N   L   M

AAT ACA ATC CAA GCT ACG GTG GCT TCG GCA GAA CGT GTC TTT GAA GTA TTA GAT
---------------------------------------------------------------------
 N   T   I   Q   A   T   V   A   S   A   E   R   V   F   E   V   L   D

GAA GAA GAA ATG GTG GAT GAA CCT TCT GGC ATA CCA GTG GAA ACA GAT AGT CCT
---------------------------------------------------------------------
 E   E   E   M   V   D   E   P   S   G   I   P   V   E   T   D   S   P

TAT CGT GTT TCT TTT GAA CAT GTT GCT TTT GGC TAT TCA CCA GAG AAA TTA TTA
---------------------------------------------------------------------
 Y   R   V   S   F   E   H   V   A   F   G   Y   S   P   E   K   L   L

ATG AAA GAT TTC AAT TTA AAT GTT AAA CCT GGG GAA ATG GTC GCA ATC GTG GGC
---------------------------------------------------------------------
 M   K   D   F   N   L   N   V   K   P   G   E   M   V   A   I   V   G

CCA ACA GGT GCT GGG AAA ACA ACC CTA ATT AAC TTA TTG GAA CGC TTC TAT GAT
---------------------------------------------------------------------
 P   T   G   A   G   K   T   T   L   I   N   L   L   E   R   F   Y   D

ATT AGT AGC GGC AGC ATC AAA TAT GAT GGC GTA GAT ACG CGC GAT TTA TCT CGC
---------------------------------------------------------------------
 I   S   S   G   S   I   K   Y   D   G   V   D   T   R   D   L   S   R

GAA GAG TTG CGA GCA CAC TTT TCA ATG GTT CTT CAA GAT ACT TGG TTG TTC ACT
---------------------------------------------------------------------
 E   E   L   R   A   H   F   S   M   V   L   Q   D   T   W   L   F   T

GGA AGT ATC TAT GAC AAT ATT CAT TAT GGT AAT GAG CAA GCG AGC GAA GAA GAA
---------------------------------------------------------------------
 G   S   I   Y   D   N   I   H   Y   G   N   E   Q   A   S   E   E   E

GTG ATC CGT GCG GCG AAA GCA GCC CAT GTG GAT GAT TTT GTC AGA AAA TTA CCA
---------------------------------------------------------------------
 V   I   R   A   A   K   A   A   H   V   D   D   F   V   R   K   L   P

GAA GGC TAT CAA ACG ATT CTA AAT GAA GAA GCC AGC AAT ATT TCT CAA GGT CAA
---------------------------------------------------------------------
 E   G   Y   Q   T   I   L   N   E   E   A   S   N   I   S   Q   G   Q

CGA CAA TTA ATT ACA ATT GCT CGA GCA TTC TTA GCA AAT CCA GAC GTT TTA ATT
---------------------------------------------------------------------
 R   Q   L   I   T   I   A   R   A   F   L   A   N   P   D   V   L   I

TTG GAT GAA GCC ACC TCA AGT GTG GAC ACT CGA ACA GAA ATA CTG ATT CAA GCA
---------------------------------------------------------------------
 L   D   E   A   T   S   S   V   D   T   R   T   E   I   L   I   Q   A

GCA ATG AAT CGT TTA TTG GAA AAT CGG ACC AGC TTT GTA GTC GCT CAT CGC TTG
---------------------------------------------------------------------
 A   M   N   R   L   L   E   N   R   T   S   F   V   V   A   H   R   L
```

Fig. 7B

```
TCG ACT ATC CGT GAT GCC GAT ACG ATT ATC GTT ATG GCA GAA GGT TCA ATT GTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   T   I   R   D   A   D   T   I   I   V   M   A   E   G   S   I   V

GAA ACA GGA ACC CAC GAT GAA TTA ATG GCG AAA AAC GGT TTT TAT GCT GAC TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   T   G   T   H   D   E   L   M   A   K   N   G   F   Y   A   D   L

TAT AAC AGT CAG TTT TCA GAA GAA GTA GCC TAA 3'
--- --- --- --- --- --- --- --- --- --- ---
 Y   N   S   Q   F   S   E   E   V   A   *
```

Fig. 7C

```
5'  ATG TCA TTA TGG TCG GTC TTT GCA GCG GTT GTA TTT ATG ATC GTA CAA ATT ATC
    M   S   L   W   S   V   F   A   A   V   V   F   M   I   V   Q   I   I

    GCA GAT TTA TAT CTG CCC ACA TTA ACA TCT AAT ATC ATT AAT GAC GGT GTC GCT
    A   D   L   Y   L   P   T   L   T   S   N   I   I   N   D   G   V   A

    AAA GGA GAT ATT AAC TAT ATT TGG AGT ACA GGA GTT GTC ATG TTA GGC TTT TCT
    K   G   D   I   N   Y   I   W   S   T   G   V   V   M   L   G   F   S

    TTG ATT AGC ATT GTC GCT TCT ATT GGA AAT ACC TTT TTT GCC ACC AAA GAA TCA
    L   I   S   I   V   A   S   I   G   N   T   F   F   A   T   K   E   S

    CAA AGT TTA GGA AAA AAA TTA CGG ACT GAT ATT TAT CGA AAA GTT GAA AAT TTT
    Q   S   L   G   K   K   L   R   T   D   I   Y   R   K   V   E   N   F

    TCT AAT AAT GAA TTT GAT AAA TTT GGG ACA GCT TCG CTA ATT ACA CGA ACA ACC
    S   N   N   E   F   D   K   F   G   T   A   S   L   I   T   R   T   T

    AAT GAT GTC AAT CAG ATT CAA ATG GTT ATG CAA ATG TTT TTA CGT TTA ATG ATT
    N   D   V   N   Q   I   Q   M   V   M   Q   M   F   L   R   L   M   I

    AAC GCA CCT TTA ACG TTA ATT GGT GCA AGT TTT ATG GCG TAT AAC AAG GAC CCA
    N   A   P   L   T   L   I   G   A   S   F   M   A   Y   N   K   D   P

    CAA TTA ACG AAA ATT TTT CTA TAT GTT TTA CCA ATT ATG GCT GTT TTA GTG GGA
    Q   L   T   K   I   F   L   Y   V   L   P   I   M   A   V   L   V   G

    GGT ATC ATG TTT TTA GCG GTT CCG TTA TTT AAA AGC ATG CAA AAG AAA ACA GAC
    G   I   M   F   L   A   V   P   L   F   K   S   M   Q   K   K   T   D

    CGT TTA AAT TTA GTT TTT CGT GAA GGA CTG ACA GGA GTT CGT GTG ATT CGG GCC
    R   L   N   L   V   F   R   E   G   L   T   G   V   R   V   I   R   A

    TTT GGC AAA GCA AAC TAT GAA GAA CAA CGT TTT GAT GAA GCC AAT AAG GAT TAT
    F   G   K   A   N   Y   E   E   Q   R   F   D   E   A   N   K   D   Y

    ACA CAA ACG GCG ATT AAA GTG AAT ACG ATT GTC GCT TTG ATG ATT CCT TTA ATG
    T   Q   T   A   I   K   V   N   T   I   V   A   L   M   I   P   L   M

    ACA CTC ATT ATG AGT GGT ACC AAT ATT GCA ATC ACC TGG TTT GGC GGT CAT TAT
    T   L   I   M   S   G   T   N   I   A   I   T   W   F   G   G   H   Y

    ATT GCG GAA ATG CAA CTA GAG GTA GGT AAC TTA ATT GCC TTT ATG ACT TAT GCC
    I   A   E   M   Q   L   E   V   G   N   L   I   A   F   M   T   Y   A

    ATG CAA ATT CTG ATG AGT TTC ATG ATG TTA TCC ATG ATT TTT GTC ATG GTA CCG
    M   Q   I   L   M   S   F   M   M   L   S   M   I   F   V   M   V   P
```

Fig. 8A

```
CGG GCG CAA GCT TCT GCT GAT CGG ATT AAT GAG GTG CTA AAT ACT GAT TCA GAA
 R   A   Q   A   S   A   D   R   I   N   E   V   L   N   T   D   S   E

ATC AAA GAT GTA CCC AAT CCA GAA CTA CTC TCT CTA AAA GGC GAC AAA GCA ACA
 I   K   D   V   P   N   P   E   L   L   S   L   K   G   D   K   A   T

TTG GCG TTT GAA CAT GTT AAT TAT CGT TAC CAG CAT GCT GAA AAC TTA GCA CTG
 L   A   F   E   H   V   N   Y   R   Y   Q   H   A   E   N   L   A   L

GAA GAT ATA GAT TTT TCA GCA AAA TCT GGG GAG ACG GTC GCG ATT ATT GGT GGG
 E   D   I   D   F   S   A   K   S   G   E   T   V   A   I   I   G   G

ACT GGT TCT GGG AAA ACA ACG CTG GTT AAT CTT TTA CCA CGC TTT TAT GAT GTT
 T   G   S   G   K   T   T   L   V   N   L   L   P   R   F   Y   D   V

GAA TCT GGC AAG ATT TTG CTA AAT GGT AAA AAC ATT AAA GAT ACG TCA CAG CAT
 E   S   G   K   I   L   L   N   G   K   N   I   K   D   T   S   Q   H

AAT TTA CGA GAA ATG ATT GGC TTT GTA CCG CAA AAG GCT GTT TTA TTT ACA GGA
 N   L   R   E   M   I   G   F   V   P   Q   K   A   V   L   F   T   G

ACG ATT CGT GAA AAT ATG CAA TAC GGC GCC CCA AAT GCG ACC GAT GAA GAA ATT
 T   I   R   E   N   M   Q   Y   G   A   P   N   A   T   D   E   E   I

TGG CAA GCC TTA GAA ATT GCG CAG GCG AAG GCG TTT GTT TCA GAA TTG GCA GAA
 W   Q   A   L   E   I   A   Q   A   K   A   F   V   S   E   L   A   E

GGT TTG GAT AGT CAT GTC GAA CAA GGC GGC GGT AAT TTC TCT GGC GGA CAA CGC
 G   L   D   S   H   V   E   Q   G   G   G   N   F   S   G   G   Q   R

CAG CGG TTG GCT ATT GCT CGT GCT TTA GTA AAA CCA GCA GAT GTC TAT GTT TTT
 Q   R   L   A   I   A   R   A   L   V   K   P   A   D   V   Y   V   F

GAT GAT TCC TTT TCT GCA TTG GAT TTC AAA ACA GAT GCT AAT TTA AGA AAA GCC
 D   D   S   F   S   A   L   D   F   K   T   D   A   N   L   R   K   A

TTA AAA GAG CAG ATG ACG GAT GCA ATT GTC GTT TTA GTA GCC CAA CGT GTA AGT
 L   K   E   Q   M   T   D   A   I   V   V   L   V   A   Q   R   V   S

ACG GTT ATG GAA GCA TCG ACA ATT TTA GTA TTA GAC GAA GGG AAA TTA GTT GGT
 T   V   M   E   A   S   T   I   L   V   L   D   E   G   K   L   V   G

AAA GGA ACG CAT GAA GAG TTA TTA GCG AAT AAC CAA ACG TAT CAA GAA ATT GTA
 K   G   T   H   E   E   L   L   A   N   N   Q   T   Y   Q   E   I   V

CAT TCT CAA TTG AGA GAG GAG GAC CTT GCA TGA 3'
 H   S   Q   L   R   E   E   D   L   A   *
```

Fig. 8B

```
5'  ATG TTA AAG CGT TTT TTT GGG TAT TAT CGG CCG TAT CGT CGA TTA TTT ATT TTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   L   K   R   F   F   G   Y   Y   R   P   Y   R   R   L   F   I   L

    GAT TTT GGG TGT GCC GTT TTT GCG GGG CTG TTA GAG TTG GCT TTT CCA GTA GCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   F   G   C   A   V   F   A   G   L   L   E   L   A   F   P   V   A

    GTC AAC CAA GTA ATT GAT AAG ATT ATG CCA AAA GGA GAT TTT CGC CTC ATT GCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   N   Q   V   I   D   K   I   M   P   K   G   D   F   R   L   I   A

    TTA GCT AGT GCG GGC TTA TTT GCA TTT TAT ATA ATC AAT ACG TTT TTA CAA TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   A   S   A   G   L   F   A   F   Y   I   I   N   T   F   L   Q   F

    ATT GTG GTT TAT TTT GGT CAT ATG TTA GGC GTG AAT ATT GAA ACA GAT ATG CGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   V   V   Y   F   G   H   M   L   G   V   N   I   E   T   D   M   R

    GAA GAA TTA TAT CAA CAT CTG CAA ACA CAA CCG TTT GAG TAT TAT GAT AAT CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   E   L   Y   Q   H   L   Q   T   Q   P   F   E   Y   Y   D   N   Q

    AAA ACA GGG AAA TTA ATG AGC CGT TTG ACC ACT GAT CTT TTT GAA ATT TCG GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   T   G   K   L   M   S   R   L   T   T   D   L   F   E   I   S   E

    GTG GCG CAT CAT GGT CCA GAA GAC GTC TTT ATT ACG ATT ATG ACG TTA GCA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   A   H   H   G   P   E   D   V   F   I   T   I   M   T   L   A   G

    TCT TTT TTA TTA ATG TTA ACC ATT CAC GTG AAA TTA GCC ATC GCT ACG TTT ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   F   L   L   M   L   T   I   H   V   K   L   A   I   A   T   F   I

    TTA TTG CCA TTT ATC ACG ATT GCG CTA GGC TAT TTT AAT AAA AAA ATG ACA AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   L   P   F   I   T   I   A   L   G   Y   F   N   K   K   M   T   K

    GCG AAT ACG GAT ATT TAC GAT AAT TTA GGT GAA TTT AAT GCT GGG ATT GAA GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   N   T   D   I   Y   D   N   L   G   E   F   N   A   G   I   E   A

    TCA GTA AGT GGG ATA CGT GTC ACG CAA TCT TTT GCA AAT GAA CCG TTT GAA CGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   V   S   G   I   R   V   T   Q   S   F   A   N   E   P   F   E   R

    AAA CAG TTT AAT TAT TTA AAT CAA ATG TAT CGA AAA TCG AAA TTA TAT TTT TAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   Q   F   N   Y   L   N   Q   M   Y   R   K   S   K   L   Y   F   Y

    AAG GTG ATG GGT GTC AGT TCC GCC TAT AAT TAC TTA TTA ATT CGT CTG ATA AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   V   M   G   V   S   S   A   Y   N   Y   L   L   I   R   L   I   N

    TTA TTT TCA TTA ATT TTT GGT GCG TAT TAT ACA ATT AAA GGA GAA ATT ACA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   F   S   L   I   F   G   A   Y   Y   T   I   K   G   E   I   T   E

    GGT CAG TTT GTA GGA TTT ATT TTA TTA GCC AAT ATT TTT ATT CGG CCG ATT GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   Q   F   V   G   F   I   L   L   A   N   I   F   I   R   P   I   E
```

Fig. 9A

```
AAA GTC AAC AAT ATG ATT GAA AGC TAT CCC AAA GGA TTT GCT GGC TTT AAA CGT
 K   V   N   N   M   I   E   S   Y   P   K   G   F   A   G   F   K   R

TTT ACG GAA GAA ATG GAC AAA CAA CCG TCC ATC AAA GAT TTG CCA GGA GCT GTT
 F   T   E   E   M   D   K   Q   P   S   I   K   D   L   P   G   A   V

GCT GTT TCA CAT TTA GAA GGA ACG ATT GCC TAT AAG GAT GTT TCA TTT GCT TAT
 A   V   S   H   L   E   G   T   I   A   Y   K   D   V   S   F   A   Y

GAA GAT GGC ACA AAA GTT TTA GAC CAT ATC AAT TTA AAA ATT CAA CCT GGT GAA
 E   D   G   T   K   V   L   D   H   I   N   L   K   I   Q   P   G   E

ACC GTT GCT TTC GTT GGA CAA AGT GGT TCA GGA AAA ACG ACC TTG TGC AAT CTG
 T   V   A   F   V   G   Q   S   G   S   G   K   T   T   L   C   N   L

TTG CCT CGT TTT TAT GAA GTC AGT AGT GGT GAA ATT ACG ATT GAC GGT AGA AAT
 L   P   R   F   Y   E   V   S   S   G   E   I   T   I   D   G   R   N

ATC CAG AAA ATG ACC TTG GCT TCA TTA CGT AAG CAG ATT GGA ATT GTT CAA CAA
 I   Q   K   M   T   L   A   S   L   R   K   Q   I   G   I   V   Q   Q

GAT GTG TTT TTA TTC CCT GGA ACA TTA CGA GAA AAT ATT GCT TAT GGG AAT TTA
 D   V   F   L   F   P   G   T   L   R   E   N   I   A   Y   G   N   L

AAT GCT ACA GAA ATA GAC ATT CAA CAA GCG GTG AAA TTA GCA CAT TTA GAA CAT
 N   A   T   E   I   D   I   Q   Q   A   V   K   L   A   H   L   E   H

GTA ATT CAG CTA ATG CCA GAT GGC TTG GAC ACA ATT ATT GGC GAA CGT GGC GTG
 V   I   Q   L   M   P   D   G   L   D   T   I   I   G   E   R   G   V

AAA TTA TCG GGC GGT CAA AAG CAG CGA GTG GCA ATT GCG CGA ATG TTC TTG AAG
 K   L   S   G   G   Q   K   Q   R   V   A   I   A   R   M   F   L   K

AAC CCG CCG ATT TTA ATT CTA GAT GAA GCG ACT TCT GCA TTA GAT ACA GAA ACA
 N   P   P   I   L   I   L   D   E   A   T   S   A   L   D   T   E   T

GAG CAA GTG ATT CAA GAG TCA TTA AAC TCG TTA GCT GAC GGG CGA ACA ACA TTA
 E   Q   V   I   Q   E   S   L   N   S   L   A   D   G   R   T   T   L

ATT ATT GCC CAC AGA CTG GCA ACG ATT AAA CAT GCT GAT CGA ATA ATC GTA GTG
 I   I   A   H   R   L   A   T   I   K   H   A   D   R   I   I   V   V

AGT GAC CAA GGA ATT TTA GAA GAT GGA ACA CAT GAA ACT TTG TAC GCG CAA AGA
 S   D   Q   G   I   L   E   D   G   T   H   E   T   L   Y   A   Q   R

GGC CAC TAT CGT CGC TTG TAC GAT GCT CAA TTT AGA ACA TAA 3'
 G   H   Y   R   R   L   Y   D   A   Q   F   R   T   *
```

Fig. 9B

```
5'  ATG ACT GAT TTA ATA AAA GCT AGT AAA TTT TTC TAT CAT TAT TTG AAA CGT TAC
    M   T   D   L   I   K   A   S   K   F   F   Y   H   Y   L   K   R   Y

    AAA GTT TCG TTC CTC TTT ATT TTC TTA GCA ATT TTC GCA GCG ACT TAT TTA CAA
    K   V   S   F   L   F   I   F   L   A   I   F   A   A   T   Y   L   Q

    GTC AAA GCG CCG CAA TTC GTT GGG GAA GCT ATT CAG GAA TTA GCG AAA TAT GCG
    V   K   A   P   Q   F   V   G   E   A   I   Q   E   L   A   K   Y   A

    GTT AAT GTG ATG CAA GGA AAA GAC GAT AAA AGT GCG TTC GTT TCT GTC ATT TGG
    V   N   V   M   Q   G   K   D   D   K   S   A   F   V   S   V   I   W

    AAA CTA CTC ATT TTT TAT GTC TTA ACT AGT GCC GCT AGT TTC ATT TAT AGT ATT
    K   L   L   I   F   Y   V   L   T   S   A   A   S   F   I   Y   S   I

    CTC TTT ACA CAA GTC GTG GGG AAA TCG ACG AAC CGC ATG CGG ATT GGT TTG TTT
    L   F   T   Q   V   V   G   K   S   T   N   R   M   R   I   G   L   F

    AAC AAA TTA GAA AAA TTG ACG ATT CGT TTC TTT GAT TCT CAT CAA GAT GGT GAA
    N   K   L   E   K   L   T   I   R   F   F   D   S   H   Q   D   G   E

    ATT TTA AGT CGT TTT ACT AGT GAC TTA GAC AAC ATC CAA AAT AGC TTA AAC CAA
    I   L   S   R   F   T   S   D   L   D   N   I   Q   N   S   L   N   Q

    GCG TTG CTA CAA GTA TTA ACC AAT ATT GCC TTA TTG GTT GGT GTC TTA ATC ATG
    A   L   L   Q   V   L   T   N   I   A   L   L   V   G   V   L   I   M

    ATG TTC CGT CAA AAT GTG GAA CTG GCA TGG GCC ACA ATT GCT TCT ACG CCG ATT
    M   F   R   Q   N   V   E   L   A   W   A   T   I   A   S   T   P   I

    GCG ATT TTA ATT GCG GTC TTT GTG ATT AGC AAG GCG CGC AAA TAT GTC GAT TTA
    A   I   L   I   A   V   F   V   I   S   K   A   R   K   Y   V   D   L

    CAG CAA GAT GAA GTG GGT AAA TTA AAT GGC TAT ATG GAT GAA AAA ATT AGT GGG
    Q   Q   D   E   V   G   K   L   N   G   Y   M   D   E   K   I   S   G

    CAA CGT GTG ATT ATC ACT AAT GGC TTA CAA GAA GAA ACC ATT GAC GGC TTT TTA
    Q   R   V   I   I   T   N   G   L   Q   E   E   T   I   D   G   F   L

    GAG CAA AAT GAA AAA GTT CGT GCC GCT ACG TAT AAA GGT CAA GTG TAT TCA GGA
    E   Q   N   E   K   V   R   A   A   T   Y   K   G   Q   V   Y   S   G

    TTA CTT TTC CCA ATG ATG CAA GGA ATG TCA TTA GTC AAT ACG GCG ATT GTT ATT
    L   L   F   P   M   M   Q   G   M   S   L   V   N   T   A   I   V   I

    TTC TTT GGT GGT TGG TTA GCA ATC AAT GGC TCT GTT GAT CGT GCC GCT GCG CTA
    F   F   G   G   W   L   A   I   N   G   S   V   D   R   A   A   A   L
```

Fig. 10A

```
GGT TTA GTC GTT ATG TTT GTC CAA TAT TCT CAA CAA TAT TAT CAA CCA TTA ATG
 G   L   V   V   M   F   V   Q   Y   S   Q   Q   Y   Y   Q   P   L   M

CAA ATT TCT TCA GGT TAT AGC ATG ATT CAA TTA GCT GTT ACA GGC GCT CGT CGT
 Q   I   S   S   G   Y   S   M   I   Q   L   A   V   T   G   A   R   R

TTA AAT GAG ATG TTT GAT GAA CCA GAT GAA ATT CGT CCC GAA AAT GGC GAA AAA
 L   N   E   M   F   D   E   P   D   E   I   R   P   E   N   G   E   K

TTA GAA GAG ATT AAC AAA GCT GTG GCT TTA AAT CAT GTT GTT TTT GGC TAT AAT
 L   E   E   I   N   K   A   V   A   L   N   H   V   V   F   G   Y   N

CCA GAA ACG CCT GTG TTG AAA GAT GTT TCT ATT CAT GTG GAT AAA GGT GAA ATG
 P   E   T   P   V   L   K   D   V   S   I   H   V   D   K   G   E   M

GTC GCT TTA GTT GGA CCG ACA GGC TCA GGA AAA ACA ACG ATT ATG AAT TTA ATG
 V   A   L   V   G   P   T   G   S   G   K   T   T   I   M   N   L   M

AAT CGC TTT TAT GAT GTG AAT GAA GGC GCC GTG ACG TTT GAT GGT GTG GAC ATT
 N   R   F   Y   D   V   N   E   G   A   V   T   F   D   G   V   D   I

CGT GAA ATG GAT TTA GAT AGC TTG CGT TCA CAT GTA GGC ATT GTT TTG CAA GAG
 R   E   M   D   L   D   S   L   R   S   H   V   G   I   V   L   Q   E

TCC GTT CTA TTT TCA GGA ACA ATT CGT GAG AAT ATT GCT TTT GGT AAG CCA GAG
 S   V   L   F   S   G   T   I   R   E   N   I   A   F   G   K   P   E

GCA ACG GAT GAA GAA ATT GTT CAA GCA GCG AAA CAA GCG AAT ATC CAT GAA TTT
 A   T   D   E   E   I   V   Q   A   A   K   Q   A   N   I   H   E   F

ATT GTA AAT CTT GAG CAA GGC TAC GAT ACT GAG ATT ACG GAA GAG AAT AAC CTT
 I   V   N   L   E   Q   G   Y   D   T   E   I   T   E   E   N   N   L

TTC AGT ACG GGG CAA AAG CAA TTA GTA AGT ATT GCA CGA ACG ATT ATT ACC AAT
 F   S   T   G   Q   K   Q   L   V   S   I   A   R   T   I   I   T   N

CCA GAA TTA TTA ATT TTG GAT GAA GCA ACA AGT AAT GTT GAT ACA GTA ACA GAA
 P   E   L   L   I   L   D   E   A   T   S   N   V   D   T   V   T   E

GCT AAA ATC CAA AAA GCG ATG GAC GAA GCA ATT AAA GGG CGC ACC AGT TTT GTG
 A   K   I   Q   K   A   M   D   E   A   I   K   G   R   T   S   F   V

ATT GCC CAT CGT TTG AAA ACG ATT CTA AAT GCG GAT CGG ATT ATT GTT TTA CGA
 I   A   H   R   L   K   T   I   L   N   A   D   R   I   I   V   L   R

GAC GGT GAA GTC ATT GAA GAA GGC AAC CAC CAT GAA CTA GTT GAA CAA GAT GGA
 D   G   E   V   I   E   E   G   N   H   H   E   L   V   E   Q   D   G
```

Fig. 10B

```
TTC TAT GCG GAA CTT TAT AAA AAT CAA TTT GTT TTT GAA TAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   Y   A   E   L   Y   K   N   Q   F   V   F   E   *
```

Fig. 10C

```
5'  ATG TTA GGC CCT TAC TTT ACT AAA ATT GCG ATT GAC CAA GTA ATT CCA CAA AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   L   G   P   Y   F   T   K   I   A   I   D   Q   V   I   P   Q   K

    AAT CTC TCG TTA CTC TTG ATA CTA GGG GCT ATT TTC CTA TTT TCC TTG GTG ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   L   S   L   L   L   I   L   G   A   I   F   L   F   S   L   V   I

    ATT GGT TGG TGT ATG CGT TAT CGG ATA TAT GCC ATC ACA GAA ATT GGA CAA GAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   G   W   C   M   R   Y   R   I   Y   A   I   T   E   I   G   Q   D

    ATT TTA AAA GAT ATG CGT TTT TCA ATT TTT GAG CAT CTA CAA AAA TTA CCA TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   L   K   D   M   R   F   S   I   F   E   H   L   Q   K   L   P   F

    TCT TAC TTT GAC AGT CGG CCG CAT GGA AAA ATC TTA ATT CGG GTG GTC AAT TAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   Y   F   D   S   R   P   H   G   K   I   L   I   R   V   V   N   Y

    ATC AAT ACG TTA AGT GAT TTA TTA AGT AAC GGA TTG ATC AAT TTA ATT TCT GAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   N   T   L   S   D   L   L   S   N   G   L   I   N   L   I   S   D

    TTA TTT AAT GTG ATT ATC ACC TTA ATC TTC ATG TTA TTT ATT GAT GTG AAA TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   F   N   V   I   I   T   L   I   F   M   L   F   I   D   V   K   L

    ACC TTA TAC AGT TTG CTG CTA TTG CCT GTG TTG TTC GTG ATG GTT CTA TTT ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   L   Y   S   L   L   L   L   P   V   L   F   V   M   V   L   F   I

    CAA GGG AAA CAA CGG AAA GCT TAC CAA GAG CTT AGC AAT AAA CAA TCA AAC TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   G   K   Q   R   K   A   Y   Q   E   L   S   N   K   Q   S   N   L

    AAT GCC TAT ATT CAT GAA AGT ATT TCA GGA ATT AAA ATT ACA CAA TCT TTT GCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   A   Y   I   H   E   S   I   S   G   I   K   I   T   Q   S   F   A

    CGA GAA GAT GAA AAT TTC CAG ATT TTT AAC GAG GTA AGC GAA GAA TAT CGT CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   E   D   E   N   F   Q   I   F   N   E   V   S   E   E   Y   R   Q

    TCC TTT ATG AAA GCT GTG CGT GTC CAG TAT TTA TTA TGG CCA GCA GTG CAA AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   F   M   K   A   V   R   V   Q   Y   L   L   W   P   A   V   Q   N

    ATA TCC GTC ATT ACG ACT TGT TTT ATC TAT TTC GTA GGA ATC CGT CAA TTA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   S   V   I   T   T   C   F   I   Y   F   V   G   I   R   Q   L   G

    GTG TCC GTT ACC ACT GGG ACG CTA ATT GCT TTT ATC GGT TAT ATC AAT AAC TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   S   V   T   T   G   T   L   I   A   F   I   G   Y   I   N   N   F

    TGG AAT CCA GTC ATC AAT ATT GGT AAC TTC TAC AAT TCT CTG ATT ACT GCA ACT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     W   N   P   V   I   N   I   G   N   F   Y   N   S   L   I   T   A   T

    GCC TAT TTG GAA AGG ATT TTT GAA ACG ATG GAT GTT GTT CCA GAA ATT CAA GAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   Y   L   E   R   I   F   E   T   M   D   V   V   P   E   I   Q   D
```

Fig. 11A

```
GCG CCG CAC GCT ATT GAG TTG CCG CCA ATT AAA GGA ACC GTC GAC TTT CAG CAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   P   H   A   I   E   L   P   P   I   K   G   T   V   D   F   Q   H

GTT TAT TTC CGT TAT GAA GAA GGA AAA AAT ATT TTG ACC GAT GTT AGT TTT CAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   Y   F   R   Y   E   E   G   K   N   I   L   T   D   V   S   F   H

ATT GAG CCA GGA CAA ACA ATC GCC TTG GTT GGG CCA ACA GGT GCG GGT AAA ACA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   E   P   G   Q   T   I   A   L   V   G   P   T   G   A   G   K   T

ACA ATC ATC AAT TTG TTA AGT CGT TTT TAT GAT GTG AAT GAA GGG GCC GTT AAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   I   I   N   L   L   S   R   F   Y   D   V   N   E   G   A   V   K

ATT GAT GGT TAT GAT GTT CGC GAT GTG ACA CTC CGT TCG TTA AGA AAA CAA ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   D   G   Y   D   V   R   D   V   T   L   R   S   L   R   K   Q   M

GGG GTA ATG CTT CAA GAT ACG TTT ATT TTT TCA GGA ACG ATT ATT GAA AAC ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   V   M   L   Q   D   T   F   I   F   S   G   T   I   I   E   N   I

CGG TAC GGA AAT TTA GCA GCC ACA GAG GAA GAA GTC ATC CAA GCC GCA AAA ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   Y   G   N   L   A   A   T   E   E   E   V   I   Q   A   A   K   I

GTT CGT GCG CAC GAC TTT ATC AAG GAT TTA AAA GAT GGC TAT GAA ACA GTT GTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   R   A   H   D   F   I   K   D   L   K   D   G   Y   E   T   V   V

GAA GAG CGG GGT AGT ACA CTC TCG GCA GGA CAA CGC CAA TTA ATT TCA TTT GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   E   R   G   S   T   L   S   A   G   Q   R   Q   L   I   S   F   A

CGT GCT TTA CTG GCA GAT CCC AAA ATT TTA ATT TTA GAC GAA GCA ACC TCC AGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   A   L   L   A   D   P   K   I   L   I   L   D   E   A   T   S   S

ATT GAT ACA AAA ACA GAA GAA TTG TTA CAA GAA GGA CTA CAA CAA CTT CTG AAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   D   T   K   T   E   E   L   L   Q   E   G   L   Q   Q   L   L   K

GGA CGG ACA TCG TTT ATT ATT GCT CAT CGA CTT TCT ACC ATC AAA AAT AGC GAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   R   T   S   F   I   I   A   H   R   L   S   T   I   K   N   S   D

AAA ATT TTC TAT ATT GAT GGT GGT CGT ATT GTG GAA GAA GGA TCT CAT GAT CAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   I   F   Y   I   D   G   G   R   I   V   E   E   G   S   H   D   Q

CTA ATG GCA AAA CAT GCG TTA TAT CAT CAT TTA TAT CAA TCG CAA TAC GAC TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   M   A   K   H   A   L   Y   H   H   L   Y   Q   S   Q   Y   D   L

TTA AAG AGC TAG 3'
--- --- --- ---
 L   K   S   *
```

Fig. 11B

```
5'  ATG CAA GAT AGT ATA CAA GAT TGT GGT GTT GCA TGC ATA GAA ATG ATA TGT AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   Q   D   S   I   Q   D   C   G   V   A   C   I   E   M   I   C   K

    TTC TAT AAT ATT AAT ATA GAT AGA AGA TAT ATC CAA GAA GAA ACT GGC TAT GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   Y   N   I   N   I   D   R   R   Y   I   Q   E   E   T   G   Y   G

    ATG ATA GGA ATA TCA TTA AAA GCT ATG GAA AAA TTT TTT TCA AAA GTA GAT GCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   I   G   I   S   L   K   A   M   E   K   F   F   S   K   V   D   A

    AAC CCA GAA ATA GTA AAT ATA TCT AAA ATT AAT AGG TTA AAC AAA GAA AAT AGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   P   E   I   V   N   I   S   K   I   N   R   L   N   K   E   N   R

    GAA ATG ATT AAT AAC TCA CTG CCA GCA ATT GTC TTT CTT GAA GAG GAG GAA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   M   I   N   N   S   L   P   A   I   V   F   L   E   E   E   E   I

    ATT AAT CAT TTT GTA GTA ATC TGG CAT ATT GGT AAG AAA AGA ATA CTT GTT TCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   N   H   F   V   V   I   W   H   I   G   K   K   R   I   L   V   S

    GAT CCT ACA CAT ACT AAA AAA GAA TGG ATA AAT AAT AAA CAT TTT GAA AAA AGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   P   T   H   T   K   K   E   W   I   N   N   K   H   F   E   K   R

    GCT ATT AGT TAT TTA TTT GTA GAA AAA CCA AAA AAT ATT TTT TTA AAT CGC AGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   I   S   Y   L   F   V   E   K   P   K   N   I   F   L   N   R   S

    CCC AAA AAA GTA AGA TTT TAT GGG AAG TTT ATA AAA AGA AAT TTA AAA ATA TTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   K   K   V   R   F   Y   G   K   F   I   K   R   N   L   K   I   F

    TCG TTA GTA ATG TTC TTT TCT ATA ATT GTT AGT TTG TTA AGT ACT TAT CTA TCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   L   V   M   F   F   S   I   I   V   S   L   L   S   T   Y   L   S

    TTT AAT TTA GGT AAA AAT ATA AAT ATT CTT GAA AAA AAT ATT CCT TTG CAA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   N   L   G   K   N   I   N   I   L   E   K   N   I   P   L   Q   E

    AAA ACT TTT CTT ATT CAA AGT ATA TCT ATA GGT TTT ATT GTA TTA GCT TTA GTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   T   F   L   I   Q   S   I   S   I   G   F   I   V   L   A   L   V

    CAA GTA GTA GTT AGC TAT TTC AAA AAT ATT CTA GTT ATA AAA GGA ACA TTT GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   V   V   V   S   Y   F   K   N   I   L   V   I   K   G   T   F   A

    TTT GAA ACA CAA TTA TTC AAA AAG TAT ATA AAT AAA ATG TTA ACT ATT CAT CCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   E   T   Q   L   F   K   K   Y   I   N   K   M   L   T   I   H   P

    AAA TAT TTT CAG TTA AAT AGT TCA GGA GAA TTT GTA TCT AGA ATT TCC GAT ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   Y   F   Q   L   N   S   S   G   E   F   V   S   R   I   S   D   I

    TTA CAA GTT AGT AAC GAT GTG ATG AAC TTT TTC ATA AGC GTA CTT CTA AAC TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   Q   V   S   N   D   V   M   N   F   F   I   S   V   L   L   N   F
```

Fig. 12A

```
TGT ATG TAT GTT TTC TCG ATA ATA ACA TTG TTT TAT TTT AGT AAG GTT ATG GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 C   M   Y   V   F   S   I   I   T   L   F   Y   F   S   K   V   M   A

ATT TAT CTT GTA GTT ATT TCT GCA GTA GTA ATA TTG TTT TTT GCA TAT ATT TAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   Y   L   V   V   I   S   A   V   V   I   L   F   F   A   Y   I   Y

CCA AAA ATC TAT CAT TTT GGG TAC ACA ATA TCT AAA TCC AAT ACT GAT TTT CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   K   I   Y   H   F   G   Y   T   I   S   K   S   N   T   D   F   Q

AAA AAA TTA ATT ACT TTT ATT GAT GGA TTA GAA GAA TTT AGG GCT TTA AAT GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   K   L   I   T   F   I   D   G   L   E   E   F   R   A   L   N   A

GAT AAA TAC TTT AGT GAG AAA GTC TTA GAT AAA ATA AAT AAA TTT ATT ACT AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   K   Y   F   S   E   K   V   L   D   K   I   N   K   F   I   T   N

AAT AAA AAA CTA CAA TTT TTT GGT ACA AAT AGT AAT GTT TTT TTA AGT TTA TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   K   K   L   Q   F   F   G   T   N   S   N   V   F   L   S   L   F

AAT ACT GTT GGG ACA GTA ATA TTG ATA GCT ATA GGC AGC TTA TTA GTA TTA AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   T   V   G   T   V   I   L   I   A   I   G   S   L   L   V   L   K

AAT AGT ATA AAT TTA GGA AAT TTA ACA GTA TTA ATT TCA TTA TCA GGA ATA GTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   S   I   N   L   G   N   L   T   V   L   I   S   L   S   G   I   V

GTA TCA TCC CTA AAT CAA TTT GGG TCG TTT CAA GCG CAA TAT GAA TCG ATG CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   S   S   L   N   Q   F   G   S   F   Q   A   Q   Y   E   S   M   Q

GTG GCT AGT CAT AGG CTT GAA TCT ATA TTA ATA AAT ATG GAA AAT GAG AAT GTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   A   S   H   R   L   E   S   I   L   I   N   M   E   N   E   N   V

TGC GGG GAA ATA ATT TTA GAC AAG AAA ATA GAA TCA ATA AGA TGT AAA CGT GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 C   G   E   I   I   L   D   K   K   I   E   S   I   R   C   K   R   V

TCC ATA AAA AAA GGA GAT ACT TTA TTA CTC GAC ACT GTT AAC TGT GAA ATC TAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   I   K   K   G   D   T   L   L   L   D   T   V   N   C   E   I   Y

AGA GGG AAA AAT CTT TCA ATA CGA GGT GAA AAT GGT TCA GGT AAA TCT ACA TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   G   K   N   L   S   I   R   G   E   N   G   S   G   K   S   T   L

ATT AAG TCT TTA GTA AGA TTA GAT GAT GAC TAT AGG GGA CAG ATT CTA ATT AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   K   S   L   V   R   L   D   D   D   Y   R   G   Q   I   L   I   N

AAT ATT GAC ATA AAA AAA ATC AAC TTG GAC TGT CTA CGA TCA AAA TTA GTG TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   I   D   I   K   K   I   N   L   D   C   L   R   S   K   L   V   F

GTA GAG CCC AAT CCT AAA TTT TTA GAA GGT ACA ATT AGA GAT AAT TTA TTA TTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   E   P   N   P   K   F   L   E   G   T   I   R   D   N   L   L   L
```

Fig. 12B

```
GGA CAT AAA GTT CCG AAT AGT ATA TTT AAT AAG TTG ATA AGA GAT TTT GAA ATA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   H   K   V   P   N   S   I   F   N   K   L   I   R   D   F   E   I

AAC AAA ATA TTA GAC GAT CTA CCT TTA GGT ATA AAC TTT CCT GGA GAA GCT GCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   K   I   L   D   D   L   P   L   G   I   N   F   P   G   E   A   A

ATA AAA TGC TTA TCT TCT GGA CAA AAA CAA AAG CTA GCC TTG TTT AGA GGT ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   K   C   L   S   S   G   Q   K   Q   K   L   A   L   F   R   G   I

TTA AAG AGA CCT GAT ATT TTA CTT GTA GAT GAA GGC TTC TCA AAT ATG GAT AAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   K   R   P   D   I   L   L   V   D   E   G   F   S   N   M   D   K

GAG TAT TTG GAT AGA ATT TTA CCT AAA TTT GAC TCA TGG GGA ATT AAA CTT ATA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   Y   L   D   R   I   L   P   K   F   D   S   W   G   I   K   L   I

GTA ATA GAT CAT TCA AAT CGT GTA ACA AAA GAT ATT GAT TAT ATT ACT ATG GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   I   D   H   S   N   R   V   T   K   D   I   D   Y   I   T   M   E

AAT TAC GAT ATC AAA AAT AGT TGG ATG TGA 3'
--- --- --- --- --- --- --- --- --- ---
 N   Y   D   I   K   N   S   W   M   *
```

Fig. 12C

```
5'  ATG GAA AAA CAG CCG ACA ACG AAA GAC GTT TGG AAA AAC GAT CAA TGG GTT CGT
    M   E   K   Q   P   T   T   K   D   V   W   K   N   D   Q   W   V   R

    CCT TTC TTA AAA CGC TAT AAA AAA ACC TTG TAT TTT GCC TTA CTT TTA GGC TTC
    P   F   L   K   R   Y   K   K   T   L   Y   F   A   L   L   L   G   F

    TTG ACG TTC TTT AGC GCA GGG GCA TTA ATG TTT ACG TCT GGT TAC TTA ATC AGT
    L   T   F   F   S   A   G   A   L   M   F   T   S   G   Y   L   I   S

    CGT GCG GCT TCC TTA CCA GAG AAT ATT TTA TTA ATT TAT ATT CCC ATT GTT TTA
    R   A   A   S   L   P   E   N   I   L   L   I   Y   I   P   I   V   L

    ACA AGA GCA TTC GGG ATT GGA CGT CCT GTT TTT CGT TAT GTG GAG CGT CTA ACG
    T   R   A   F   G   I   G   R   P   V   F   R   Y   V   E   R   L   T

    AGT CAT AAC TGG GTG TTA AAA ATG ACC TCT GAT CTG CGC TTG AAA CTT TAT AAT
    S   H   N   W   V   L   K   M   T   S   D   L   R   L   K   L   Y   N

    GTA CTA GAA AAA GAT GCG ATT TTC TTT AAG ACA AAA TAT CGC ACT GGG GAT ATC
    V   L   E   K   D   A   I   F   F   K   T   K   Y   R   T   G   D   I

    TTA GGT TTG CTT TCG GAG GAT ATT AAT CAT ATC CAA AAC TTG TAT TTA CGA ACG
    L   G   L   L   S   E   D   I   N   H   I   Q   N   L   Y   L   R   T

    ATT TTC CCA ACC GTT ATT GCT TGG ATT TTG TAT ATC TTT TTA GTG ATT GCA TTA
    I   F   P   T   V   I   A   W   I   L   Y   I   F   L   V   I   A   L

    GGT TTC TTT TCT TGG TGG TTT GCT TTA TGC ATG TTA CTG ATG TTA GGA GTG GTT
    G   F   F   S   W   W   F   A   L   C   M   L   L   M   L   G   V   V

    GTT TTT CTA TTA CCA CTT GTT TCA GTT CTA GTG AAT GGT GCG CGT CAA GAA AAA
    V   F   L   L   P   L   V   S   V   L   V   N   G   A   R   Q   E   K

    CAT AAA TAT GCC AAA AAT GAG TTA TAT CAA ACA CTA ACC GAC AAT ATT CTA GGT
    H   K   Y   A   K   N   E   L   Y   Q   T   L   T   D   N   I   L   G

    GTG TCT GAT TGG GTA TTT AGT CAA CGA GGC TCT GAA TTC GTT GCT CGT TAT GAA
    V   S   D   W   V   F   S   Q   R   G   S   E   F   V   A   R   Y   E

    ACA GAT GAA GCA AAC GTT CGT GCA TTA GAT GAA AAA ATG AAG CAA TTT AAC CGT
    T   D   E   A   N   V   R   A   L   D   E   K   M   K   Q   F   N   R

    GGC CGA GAC TTT GTT TTA CAA CTT CTG TTT GGT GTG ATT GCG ATT GCT GTC TTA
    G   R   D   F   V   L   Q   L   L   F   G   V   I   A   I   A   V   L

    GCT TGG ACA AGT GTG CGG TTT CCA GGT AAT CAT GGG GGC GCA GCC AAT TGG ATT
    A   W   T   S   V   R   F   P   G   N   H   G   G   A   A   N   W   I
```

Fig. 13A

```
GGT GCC TTT GTT TTA ACC GTT TTT CCG TTA ATT GAT GCC TTT GCT CCG CTA CCT
 G   A   F   V   L   T   V   F   P   L   I   D   A   F   A   P   L   P

GCT GCA GCA CAG GAA ACA ACG ATT TAT AAG GAT TCC ATT CGT CGT TTT AAT GAG
 A   A   A   Q   E   T   T   I   Y   K   D   S   I   R   R   F   N   E

CTT CCT GAA GGG GAA GAT GAC TCA ACA GAA GCG CCT GTT CAG CCA AAT GGC ACC
 L   P   E   G   E   D   D   S   T   E   A   P   V   Q   P   N   G   T

GGT TTA TCG ATA GAA CAT TTG TCT TTT GCA TAT GAA AAT CAA GCG AAA AAG GTA
 G   L   S   I   E   H   L   S   F   A   Y   E   N   Q   A   K   K   V

TTA AAT GAC TTA TCG TTA ACA ATT CCT GAA AAA CAA AAA TTA GCG ATT TTA GGG
 L   N   D   L   S   L   T   I   P   E   K   Q   K   L   A   I   L   G

CGC AGT GGT TCC GGA AAG AGT ACG TTG GCT TCG CTC ATT CGT GGT GAT TTA CGG
 R   S   G   S   G   K   S   T   L   A   S   L   I   R   G   D   L   R

CCT ACT TCG GGA GAA ATC CTT TTA GGA AAT ATC CCA ACT GAG GCA TTT GGT GAG
 P   T   S   G   E   I   L   L   G   N   I   P   T   E   A   F   G   E

ACG ATG ACT GAA TAT ATT GGG GTT ATG CAT CAA GCA CCG TAC CTA TTT CGA ACA
 T   M   T   E   Y   I   G   V   M   H   Q   A   P   Y   L   F   R   T

ACC ATT TTA AAT AAT ATT CGG ATT GGT AGA GAA GAA GCA AGC GAA GCA GAA GTT
 T   I   L   N   N   I   R   I   G   R   E   E   A   S   E   A   E   V

TGG GCG GTT TTA GAA AAA GTT GGC TTA AAA GAA ATG GTT GCC CAA TTA CCA GAA
 W   A   V   L   E   K   V   G   L   K   E   M   V   A   Q   L   P   E

GGT TTA CAA ACA ATG GTG GAT GAA GCA GGT TTG CGT TTT TCT GGT GGC GAA CGA
 G   L   Q   T   M   V   D   E   A   G   L   R   F   S   G   G   E   R

CAT CGC TTG GCG CTA GCG CGG ATT CTT TTG CAA GAT ACA CCG ATT GTG TTA TTA
 H   R   L   A   L   A   R   I   L   L   Q   D   T   P   I   V   L   L

GAC GAA CCA ACT ACT GGT TTA GAT CCA ATT ACA GAA CAG CAA TTG CTG GAA ACC
 D   E   P   T   T   G   L   D   P   I   T   E   Q   Q   L   L   E   T

TTT TTT GAG GCA TTA AAA GAT AAA ACC GTC ATC TGG ATT ACC CAC CAT TTG CAA
 F   F   E   A   L   K   D   K   T   V   I   W   I   T   H   H   L   Q

GGC GTA ACT TTG ATG GAT CAA GTG ATC TTT ATT GAA GAT GGC CAA TTA GAG ATG
 G   V   T   L   M   D   Q   V   I   F   I   E   D   G   Q   L   E   M

AGT GGT ACG CCA GAA GAA TTA CTT GCA ACC AAT GCG CAT TAT CAG AAA CTT TAC
 S   G   T   P   E   E   L   L   A   T   N   A   H   Y   Q   K   L   Y
```

Fig. 13B

```
CGA ATT GAT CGT GGC ATT AGT TCT TTC GAG GAA TAA 3'
--- --- --- --- --- --- --- --- --- --- --- ---
 R   I   D   R   G   I   S   S   F   E   E   *
```

Fig. 13C

```
5'  ATG AAA GAG TTT ATT AAA GAA AAT AAA TGG ATT GTT CTT GCG ACA ACT TTA ACC
    M   K   E   F   I   K   E   N   K   W   I   V   L   A   T   T   L   T

    ATC TGT TTG CAA ATC GCA GGA ACC CTT GGC GTC CCT AAA TTA GTT GGC AAG TTG
    I   C   L   Q   I   A   G   T   L   G   V   P   K   L   V   G   K   L

    ATT GAT GTG GGC ATC GTT AGC GGT GAC CAA CAA GTA ATT AAM ACG ATT GGC ATA
    I   D   V   G   I   V   S   G   D   Q   Q   V   I   X   T   I   G   I

    CAA ATG TTT CTT GTG GCA TTC ATT GGA ACA ATT GCC GCA ATT ATT TCT AGC TAT
    Q   M   F   L   V   A   F   I   G   T   I   A   A   I   I   S   S   Y

    TTG TCT GCT TTA GTA GCT GCT AAA TTT GGT TTT CAA GTT AGA GGA TTG TTC TTT
    L   S   A   L   V   A   A   K   F   G   F   Q   V   R   G   L   F   F

    AAA AAA TTT CAA CAA TTC TCG ATG AAA AAT GTT GAT AAA TTT GGT TCA AAT TCT
    K   K   F   Q   Q   F   S   M   K   N   V   D   K   F   G   S   N   S

    TTG CTA ACT AGA ATG ACC AAC GAT GTA GAT AAT GTT CAA ACA ATG ATT GTG CTG
    L   L   T   R   M   T   N   D   V   D   N   V   Q   T   M   I   V   L

    TTT TGC CAA TTA ATC TTT CCA GCG CCT ATT ATT AGT TTA TTT GCC TTA GTG ATG
    F   C   Q   L   I   F   P   A   P   I   I   S   L   F   A   L   V   M

    ACA TTT TCT TAT TCA GTT TCA CTC GCT TGG GTA ACA TTG GCT TCC ATC GTA TTT
    T   F   S   Y   S   V   S   L   A   W   V   T   L   A   S   I   V   F

    TAC TTA GTC GTT GTT TAT TTT TTA ATG AAA AAA GGT ACC CCT TTA TCT TTA AAA
    Y   L   V   V   V   Y   F   L   M   K   K   G   T   P   L   S   L   K

    ATT CAA CCA AAA ATG GAT CGA ATT ACT ACG ACT TTA CGA GAG TTC TTT ACT GGA
    I   Q   P   K   M   D   R   I   T   T   T   L   R   E   F   F   T   G

    ATT AAT ATG ATT CGT GCG TTC AAT AAT CAA GAT TTT GAA GAG CAG CGA ACC AAT
    I   N   M   I   R   A   F   N   N   Q   D   F   E   E   Q   R   T   N

    CAA ACA TTT AAA AAT TAT GCT GAA CGC ATG AGT AAA GTG AAT CAA ATC TTT GCT
    Q   T   F   K   N   Y   A   E   R   M   S   K   V   N   Q   I   F   A

    TGG ATT ACT CCC GTT GCC TTT TTA TTA ATG GGC GTT GTG TAC GCC TCT ATT TTG
    W   I   T   P   V   A   F   L   L   M   G   V   V   Y   A   S   I   L

    TGG TTT GGC GGT AAT TTA GTT GCA GTA GGC ACC CTA CAA ATT GGC ACC GTT ACA
    W   F   G   G   N   L   V   A   V   G   T   L   Q   I   G   T   V   T

    GCT GTG ATT GAA TAT ACG CTT CTA ACT TTG GCC TAC TTA ATG ATC GCG GCT ATG
    A   V   I   E   Y   T   L   L   T   L   A   Y   L   M   I   A   A   M
```

Fig. 14A

```
GTA TTA GTT GTC ATT CCA CGG TCC GTT GCT TCC TTG AAT CGC TTG CAA GAA GTT
 V   L   V   V   I   P   R   S   V   A   S   L   N   R   L   Q   E   V

TTG TCA GAA GAA ATT GAA ATT AGC GAT CCT CAT ACT GAG GCA ACC ATT GCT TAT
 L   S   E   E   I   E   I   S   D   P   H   T   E   A   T   I   A   Y

CAT CCT GAG AAA GCC TTG ATT TGC TTC GAT CAC GTC ACG TTT CAA TAC ACA GAA
 H   P   E   K   A   L   I   C   F   D   H   V   T   F   Q   Y   T   E

ACA GCT GAT CCT GTT TTA GAA AAT GTT AGT TTT GTC ATT CCT AAA GGA AAA ACA
 T   A   D   P   V   L   E   N   V   S   F   V   I   P   K   G   K   T

ACG GCG ATT GTT GGT GCA ACT GGC GCT GGT AAA AGT ACT TTA GTT AAG TTA CTT
 T   A   I   V   G   A   T   G   A   G   K   S   T   L   V   K   L   L

TTA CGA ATA AAT GAG GTC ACA GCC GGC ACG ATT AGC TAT TCT GGC ACA GAT ATC
 L   R   I   N   E   V   T   A   G   T   I   S   Y   S   G   T   D   I

CGC TCA TTA TCT CAG CAA ACG ATT CGC CAA GTC ATC AGT TAT GTG CCA CAA AAA
 R   S   L   S   Q   Q   T   I   R   Q   V   I   S   Y   V   P   Q   K

GCC TTT CTT TTC AGT GGG ACA ATC TTA TCA AAC TTA TTA ATG GGA AAT GCC AAA
 A   F   L   F   S   G   T   I   L   S   N   L   L   M   G   N   A   K

GCA ACT ACA GAA GAA ATA AGA ACG GCA CTA GAA ATT TCA CAA TCT TCT GAA TTT
 A   T   T   E   E   I   R   T   A   L   E   I   S   Q   S   S   E   F

ATC GAT TCC TTA CCA CAA GGG ATT GAA AGT TTC GTA GCA CAA GGC GGG TCC AAC
 I   D   S   L   P   Q   G   I   E   S   F   V   A   Q   G   G   S   N

TAT TCT GGT GGT CAA AAA CAA AGA ATG TGT ATT GCA CGA GCC TTA ATC AAA CCG
 Y   S   G   G   Q   K   Q   R   M   C   I   A   R   A   L   I   K   P

GCA GAC GTT TAT ATT TTC GAT GAT AGC TTT TCC GCA TTA GAT TAC AAA ACT GAT
 A   D   V   Y   I   F   D   D   S   F   S   A   L   D   Y   K   T   D

GCC GCT CTA CGT GCC GCT TTA CAT GCA CAA ATG TCG GAC AAA ACT TTA CTC ATT
 A   A   L   R   A   A   L   H   A   Q   M   S   D   K   T   L   L   I

GTT GCT CAA CGG TTA AGT ACA ATC ATG AAC GCT GAC AAC ATT ATC GTC CTA GAT
 V   A   Q   R   L   S   T   I   M   N   A   D   N   I   I   V   L   D

GAA GGA AGG ATT GTT GGT CAA GGC ACC CAC GCT GAT TTA CTT ACC ACT AAT AGC
 E   G   R   I   V   G   Q   G   T   H   A   D   L   L   T   T   N   S

TAT TAC CAA GAC TTT GCT AAA TCG CAA GGT ATC TTA CCC AAG TAA 3'
 Y   Y   Q   D   F   A   K   S   Q   G   I   L   P   K   *
```

Fig. 14B

```
5'  ATG AAA AAT TTT TAT AAA AAA AAA TTT GCT TTA ACT GAT CAG GGC GCA GAA GCT
    M   K   N   F   Y   K   K   K   F   A   L   T   D   Q   G   A   E   A

    TTA ACT AAA GCT TCT ATT TCT AGT TTT TTT GTA TAC TGC ATT AAT ATG GTA CCA
    L   T   K   A   S   I   S   S   F   F   V   Y   C   I   N   M   V   P

    GCA TTC ATT ATA ATG ATG CTG ATA GAC GAA TTA GTT TTA GAA AAT GCA AAG CCC
    A   F   I   I   M   M   L   I   D   E   L   V   L   E   N   A   K   P

    CGT TGG CTT TAT TTT GCA GTA TCG TTT GTT ACT TTG CTT TTC ATG TAT TGG TTG
    R   W   L   Y   F   A   V   S   F   V   T   L   L   F   M   Y   W   L

    TTA GAT AGG GAG TAT GAA AAC TTA TAT AAT AGT ACT TAT AAA GAA AGT GCG CAT
    L   D   R   E   Y   E   N   L   Y   N   S   T   Y   K   E   S   A   H

    TTA AGA GTG CAA ATT GCA GAC GAT TTG TCA AAT TTA CCA TTA TCC TAT TTT TCA
    L   R   V   Q   I   A   D   D   L   S   N   L   P   L   S   Y   F   S

    AAA CAT AAT TTA TCA GAT TTA TCT CAA ACT ATC ATG TCT GAC GTT GAA GGT ATT
    K   H   N   L   S   D   L   S   Q   T   I   M   S   D   V   E   G   I

    GAG CAT GCG ATG AGT CAT GCA ATA CCT AAA TCC GGC GGT ATG GCT CTG TTT TTC
    E   H   A   M   S   H   A   I   P   K   S   G   G   M   A   L   F   F

    CCT TTT ATT TCA GTG ATG CTT TTG GTT GGT AAT GTC AAA ATG GGA TTA GCT GTT
    P   F   I   S   V   M   L   L   V   G   N   V   K   M   G   L   A   V

    ATT TTG CCA ACG TTA TTT AGT TTT GTC TTA ATC TTG TTA TCA AAG AAA TCC CAA
    I   L   P   T   L   F   S   F   V   L   I   L   L   S   K   K   S   Q

    ACG AAA GCC AAT ACT AAA TAT TAC GAT ACT TTG AGA GAA AAC TCG GAA GAA TTT
    T   K   A   N   T   K   Y   Y   D   T   L   R   E   N   S   E   E   F

    CAA GAA ACT ATT GAA CTG CAG CAA GAG ATT AAT AGC TTT AAT CTA TCT AAA AAA
    Q   E   T   I   E   L   Q   Q   E   I   N   S   F   N   L   S   K   K

    GTT CAA GAC AGA CTT TTC AAA AAA ATG GAA GAG AGT GAA AGG ATT CAT TTA AAG
    V   Q   D   R   L   F   K   K   M   E   E   S   E   R   I   H   L   K

    GTA GAA TTA AGT ACT TTT TCA GTC ATG GCC TTA TCC TCT ATT TTC TCA TAT GTT
    V   E   L   S   T   F   S   V   M   A   L   S   S   I   F   S   Y   V

    AGT TTA GCA GTA GTC ATT CTA GTA GGT GTT CAC TTA CTG TTA ACG GGT GAA GTG
    S   L   A   V   V   I   L   V   G   V   H   L   L   L   T   G   E   V

    ACT ATA CTC TAC GTC GTT GGT TAC TTA CTA GCC GCA ATA AAA ATA AAG GAT TCC
    T   I   L   Y   V   V   G   Y   L   L   A   A   I   K   I   K   D   S
```

Fig. 15A

```
TTC GAT TCT ATG AAA GAA GCA GTG CTT GAA ATA TTT TAT TTA GCT CCT AAA ATA
 F   D   S   M   K   E   A   V   L   E   I   F   Y   L   A   P   K   I

CAG CGA ATT AGA GCT ATG AAG GAA ACC AGT ATC CAA GAG GGA TCA GAT AGT CCT
 Q   R   I   R   A   M   K   E   T   S   I   Q   E   G   S   D   S   P

CTG AAA TCT TTT GAC GTT GAG TTG CGA GAT GTC TCT TTC TCA TAT GAT AAC AAT
 L   K   S   F   D   V   E   L   R   D   V   S   F   S   Y   D   N   N

ACC CCA ATT TTA GAT CAT ATC TCT TTT ACG GCT AAA CAA GGA GAA GTG ACT GCG
 T   P   I   L   D   H   I   S   F   T   A   K   Q   G   E   V   T   A

TTA GTA GGT GCT AGT GGT TCC GGG AAA ACT AGT ATC TTG AAA TTA GTG TCT AGA
 L   V   G   A   S   G   S   G   K   T   S   I   L   K   L   V   S   R

CTG TAT GAT TAT GAT GAA GGC TGT ATT TTA ATT GAT GGG TAT GAT ATT AAA CGT
 L   Y   D   Y   D   E   G   C   I   L   I   D   G   Y   D   I   K   R

GTA TCT CCA GCT TCT TTG TTT AGC AAA ATC GCT ATT GTT TTT CAA GAG GTA ACT
 V   S   P   A   S   L   F   S   K   I   A   I   V   F   Q   E   V   T

TTG TTT AAT ACA AGT ATT TTG GAA AAT ATA CGT ATT GGA AAT AGT CAA GCG TCA
 L   F   N   T   S   I   L   E   N   I   R   I   G   N   S   Q   A   S

GAT GAG GAA GTG AAA AAA GCA GCT CGT TTA GCT AAC TGC GAA GAC TTT ATT GAA
 D   E   E   V   K   K   A   A   R   L   A   N   C   E   D   F   I   E

AAA CTA CCT GAC GGT TAT CAT ACA TTG GTC GGA GAA AAC GGA AGT TCA CTT TCT
 K   L   P   D   G   Y   H   T   L   V   G   E   N   G   S   S   L   S

GGA GGA GAG CGA CAA CGC TTA TCC ATA GCA CGG GCT TTT TTG AAA AAT GCG CCA
 G   G   E   R   Q   R   L   S   I   A   R   A   F   L   K   N   A   P

ATT TTA ATT TTA GAT GAA ATT ACT GCC AGT TTA GAT GCG GAA AAT GAA AAA CGT
 I   L   I   L   D   E   I   T   A   S   L   D   A   E   N   E   K   R

ATT CAG GAA AGT TTA AAT CGT TTA ATA CAA GAT AAG ACA GTA CTA ATT ATA TCG
 I   Q   E   S   L   N   R   L   I   Q   D   K   T   V   L   I   I   S

CAT CGC TTG AAG TCT ATT GAA AAG GTA AAT AAA ATT GTA GTG ATG GAT CAA GGA
 H   R   L   K   S   I   E   K   V   N   K   I   V   V   M   D   Q   G

AAA GTA GTT GAT CAA GGT ACT CAT AGT GAG TTG TAT CGA AGG TCA GAA ATA TAC
 K   V   V   D   Q   G   T   H   S   E   L   Y   R   R   S   E   I   Y

AAA AAT CTC ATA AAG AAA ACT AAA CTA TCT GAG AAA TTC GTC TAT GAA AAA GAA
 K   N   L   I   K   K   T   K   L   S   E   K   F   V   Y   E   K   E
```

Fig. 15B

```
GCT CAA TCA CGA TGA 3'
--- --- --- --- ---
 A   Q   S   R   *
```

Fig. 15C

```
5'  ATG ACT AAA AAA ATA TGG TCG CAC TTT CGG AAA AGC TTC ACT TGG TAC ATT TTG
    M   T   K   K   I   W   S   H   F   R   K   S   F   T   W   Y   I   L

    ATT GGC GTG ATT GTT TCA TTA TGT AGT GCT TTA GCA ATT TAT TTT TTT CAA CAG
    I   G   V   I   V   S   L   C   S   A   L   A   I   Y   F   F   Q   Q

    TTG TTA GAT CAT TAT CAA AAA AGT TTC CAA CTA GGG TTG TTA GTA GCG TAT GGA
    L   L   D   H   Y   Q   K   S   F   Q   L   G   L   L   V   A   Y   G

    ACG ACA ATT ATT TTA ATT CCT CTA TTG TCT TAC TGT GAA CAG AAA CCG AAA GCT
    T   T   I   I   L   I   P   L   L   S   Y   C   E   Q   K   P   K   A

    TAC TTA ACT AAT GGC ATC TAT TTC TAT TTG AAG AAA TTA AGT TTA ATA AAA ATG
    Y   L   T   N   G   I   Y   F   Y   L   K   K   L   S   L   I   K   M

    AGC AAG ATT TCC TAT GAA GAA TAT CTA AAA CTG GGA GCG GGG GCT TTA CTA CAA
    S   K   I   S   Y   E   E   Y   L   K   L   G   A   G   A   L   L   Q

    AAA GTA GAA GTT GGT GCA GCA GCT GGG AGA AAT ATC CAC TTG AAT TTT TAT GGG
    K   V   E   V   G   A   A   A   G   R   N   I   H   L   N   F   Y   G

    CGC TTA TTT CGA GAG TTA ATT CCA GAG ACG TTA TTT AAT TTA TTT TTC ATA GCA
    R   L   F   R   E   L   I   P   E   T   L   F   N   L   F   F   I   A

    TTG ATT GAT AAA AAA TTA TTA CCA GCA ATT CTA ATT GGT TAT GTG ATT GTA TTT
    L   I   D   K   K   L   L   P   A   I   L   I   G   Y   V   I   V   F

    ATT TTA ACC AAA ATA CTT TTG AAA ACG CTA CAA AAA ATG AAA GAA AAA ACA CTT
    I   L   T   K   I   L   L   K   T   L   Q   K   M   K   E   K   T   L

    ATT TCT GAA GAA GCA ATG AAT GCT ACT TTG ATA CGA GGG ATG ACC GAA CTA GTA
    I   S   E   E   A   M   N   A   T   L   I   R   G   M   T   E   L   V

    ACT TTT AGA ATT AAT CGA AAA TAT AAA AAA GAG ATT GAA AAT TAC GCG TTA ATG
    T   F   R   I   N   R   K   Y   K   K   E   I   E   N   Y   A   L   M

    GCA GAG GAA AAT AGT CAA AAT ATA ACA AAA ATG ACT ATG ATC CAT GAA TTT TTC
    A   E   E   N   S   Q   N   I   T   K   M   T   M   I   H   E   F   F

    TTT GGC TTT TTT GCC TTG TTA GTT GCA CTT ATC AAA GTA AGT ATT GTC GTA CTT
    F   G   F   F   A   L   L   V   A   L   I   K   V   S   I   V   V   L

    AGT TTT ACT AAC GTC GTG ACG CTA AGT TTA GGT GGA CTT GTA GCG ATA GTG ATG
    S   F   T   N   V   V   T   L   S   L   G   G   L   V   A   I   V   M

    TAT ATC GAT CGA ATT TAT ACG CCG ATA GCC ATA TTT AAT GTT TTA TTT GTT CAA
    Y   I   D   R   I   Y   T   P   I   A   I   F   N   V   L   F   V   Q
```

Fig. 16A

```
TAT AAT TTG GAT AAA GTA GCC TAT CAG CGC TTG GAA GAT TTT TAT AAA AAA GAA
 Y   N   L   D   K   V   A   Y   Q   R   L   E   D   F   Y   K   K   E

GAT GAT CCT GAC TTA GCG GTT TCC GGC AAA GCG TTA CCT GAA ATT CAA ACA ATT
 D   D   P   D   L   A   V   S   G   K   A   L   P   E   I   Q   T   I

TCA TTA AAA GAT GTG TGT TTT AGC GTA GAT AGT CAA ACG ATT ATG TCC CAA CAA
 S   L   K   D   V   C   F   S   V   D   S   Q   T   I   M   S   Q   Q

AAT CGA CAA TTT TCA ATG AAT AAA ACC TAT GGC TTG ATT GGA AAA AGT GGG ACA
 N   R   Q   F   S   M   N   K   T   Y   G   L   I   G   K   S   G   T

GGT AAA TCT ACC CTG ATT AAA TTG ATT TTA GGA TTA TTG AAG CCG ACA GAG GGG
 G   K   S   T   L   I   K   L   I   L   G   L   L   K   P   T   E   G

ACA GTT TAT GTG AAT CAA TTT CCT CTA ACA CAA TTT AAT TTG GAA GAT TAT TAT
 T   V   Y   V   N   Q   F   P   L   T   Q   F   N   L   E   D   Y   Y

GAA AAA GTT TTT TAT TTA TCT CAA GAT GTG CCA ATA TTT CAA GGA ACG TTA AAA
 E   K   V   F   Y   L   S   Q   D   V   P   I   F   Q   G   T   L   K

GAA AAT ATT GTA TTT AAT CAA GAA ATT TCT GAT GAA CAA GTA ATT GAA GCT ATG
 E   N   I   V   F   N   Q   E   I   S   D   E   Q   V   I   E   A   M

TAT CGT TTT CAG TTG GGG GAA CTT TAT GAA CGG TTG CCG GAA GGT TTA AAC ACT
 Y   R   F   Q   L   G   E   L   Y   E   R   L   P   E   G   L   N   T

ATT GTG AGT GAA AAA GGA ATG AAT TTT TCT GGT GGT GAA AAA CAA CGG ATT GCT
 I   V   S   E   K   G   M   N   F   S   G   G   E   K   Q   R   I   A

TTC ACA CGC TTA GCA TTT ACA CAA GCA GAG ATC CTT ATT TTA GAT GAG GCG ACT
 F   T   R   L   A   F   T   Q   A   E   I   L   I   L   D   E   A   T

TCT GCA TTG GAC GAA AAG ACC GAA GAA AAA GTG TTA CAA GAA GTA CAA AAA TTT
 S   A   L   D   E   K   T   E   E   K   V   L   Q   E   V   Q   K   F

ACC CAC AAT AAG TTG ACT ATT TTG GTT ACT CAT CGA CCT AAA ACC CTG CGA TTT
 T   H   N   K   L   T   I   L   V   T   H   R   P   K   T   L   R   F

GTA GAT GAA ATT ATT GAT TTA AAT GAG TGA 3'
 V   D   E   I   I   D   L   N   E   *
```

Fig. 16B

```
5'  ATG TCC ATA TTC AAA AAA TTA GGC TGG TTT TTT AAG CAA GAG AAA AAA AGT TAT
    M   S   I   F   K   K   L   G   W   F   F   K   Q   E   K   K   S   Y

    ATT ATT GGG GTT TTC TCA TTA ATG ATG GTC GCT CTT GTT CAA TTA GTC CCG CCC
    I   I   G   V   F   S   L   M   M   V   A   L   V   Q   L   V   P   P

    AAA GTT ATT GGC GTC GTT GTA GAT GAA ATC GTT AAC AAA GAA ATT CGC TTA ACG
    K   V   I   G   V   V   V   D   E   I   V   N   K   E   I   R   L   T

    AAA ATT ATC GTG TGG GTT GCA CTC TTG ATT GGT GCT GGG CTT GCC CAA TAT CTT
    K   I   I   V   W   V   A   L   L   I   G   A   G   L   A   Q   Y   L

    TTT CGC TAT ATT TGG CGG ATG CAT ATT TGG GGG AGT GCG GCT CGT TTG GAA AAA
    F   R   Y   I   W   R   M   H   I   W   G   S   A   A   R   L   E   K

    GAG CTA CGG ACT CAA TTA TTT CAT CAT TTC ACA AAA ATG GAT AGC ATC TTT TAT
    E   L   R   T   Q   L   F   H   H   F   T   K   M   D   S   I   F   Y

    CAG AAA TAT CGG ACA GGT GAC TTG ATG GCG CAT GCA ACC AAT GAT TTA AAT GCC
    Q   K   Y   R   T   G   D   L   M   A   H   A   T   N   D   L   N   A

    ATC CAA AAT GTT GCT GGA GCT GGG ATT TTA ACG TTT GCC GAC TCT GTG ATT ACG
    I   Q   N   V   A   G   A   G   I   L   T   F   A   D   S   V   I   T

    GGA GGA ACA ACG ATT ATC GCA ATG GTT CTA TTT GTC GAT TGG CGC TTA ACA TTA
    G   G   T   T   I   I   A   M   V   L   F   V   D   W   R   L   T   L

    ATT GCT TTA TTA CCG CTG CCT TTA TTA GCC GTC ACT TCA CGA GTT CTA GGT TCT
    I   A   L   L   P   L   P   L   L   A   V   T   S   R   V   L   G   S

    AAG TTG CAT GAT GCC TTT CGA GAT TCA CAA GCT GCT TTT TCA GCG ATT AAT GAT
    K   L   H   D   A   F   R   D   S   Q   A   A   F   S   A   I   N   D

    AAA ACA CAA GAA AGT ATT ACT GGA ATT AAA GTC ATT AAA ACA TTT GGG CAA GAA
    K   T   Q   E   S   I   T   G   I   K   V   I   K   T   F   G   Q   E

    AAA GAA GAT CTT GCT GAT TTT ACA GAG AAA ATT GAT GAT GCT ATC GTC AAA AAT
    K   E   D   L   A   D   F   T   E   K   I   D   D   A   I   V   K   N

    AAA CGA ACG AAC TTT TTA GAT GCA CTA TTT GAT CCG TTT ATT ACG CTG ATT ATT
    K   R   T   N   F   L   D   A   L   F   D   P   F   I   T   L   I   I

    GGT GTT TCT TAT GTT TTG ACA ATT ATT ATT GGT GGT CGT TTC ATC ATG GAA GGA
    G   V   S   Y   V   L   T   I   I   I   G   G   R   F   I   M   E   G

    ACG ATT AGT TTA GGT CAG TTG GTC TCT TTC ATT GCC TAT ATT GGG ATG CTA GTT
    T   I   S   L   G   Q   L   V   S   F   I   A   Y   I   G   M   L   V
```

Fig. 17A

```
TGG CCA ATG TTT GCG ATT GGG CGC TTA TTT AAT GTG TTA GAA CGT GGG AAT GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 W   P   M   F   A   I   G   R   L   F   N   V   L   E   R   G   N   A

AGT TAC GAT CGT GTG AAC GAA TTA TTA CAT GAA AAA ACG CAT ATT ATC GAA CGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   Y   D   R   V   N   E   L   L   H   E   K   T   H   I   I   E   R

AAA GAT GCC ATC AAA ACA ATG GCG CAA GGG ACA ATT TCA ATG AAG ATT GAT TCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   D   A   I   K   T   M   A   Q   G   T   I   S   M   K   I   D   S

TTT TCT TAT CCA AAA GAA GAG ACG GTG GCG TTG GAA AAT ATC CAA TTT TCG TTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   S   Y   P   K   E   E   T   V   A   L   E   N   I   Q   F   S   L

CAA GAA GGA GAA ACC TTG GGG ATT GTT GGC AAA ACA GGC GCT GGT AAA ACC ACT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   E   G   E   T   L   G   I   V   G   K   T   G   A   G   K   T   T

ATT TTG AAA TTG TTG ATG CGT GAA TAT GAC CAA TAC CAA GGA ACG ATT TCC TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   L   K   L   L   M   R   E   Y   D   Q   Y   Q   G   T   I   S   F

GGA AAA CAT AAC ATT AAA AAT TAC ACA TTA GAT GCA TTG ATG GGT GCA ATG GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   K   H   N   I   K   N   Y   T   L   D   A   L   M   G   A   M   G

TAC GTG CCA CAG GAC CAT TTT CTC TTT TCC ATG ACG GTA CGA GAT AAC ATT CGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Y   V   P   Q   D   H   F   L   F   S   M   T   V   R   D   N   I   R

TTT GCT AAA CCA CAC TTG GAA CAA GCA GCA GTT GAA CAA GCG GCA GCA TTA GCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   A   K   P   H   L   E   Q   A   A   V   E   Q   A   A   A   L   A

TTT ATT AAC CAA GAA ATT AAA GCA TTC CCT GAA GGC TAT GAC ACA ATG GTT GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   I   N   Q   E   I   K   A   F   P   E   G   Y   D   T   M   V   G

GAA CGT G 3'
--- --- -
 E   R
```

Fig. 17B

```
5'  ATG AAA AAG TAC GTG AGT TCG CTA AAA AGT TCG TTT CAT TTC GTG TAT AAT AAG
    M   K   K   Y   V   S   S   L   K   S   S   F   H   F   V   Y   N   K

    AGA AAG AAA AAA ATC CAT TTA ATT AAG GGA GTA ATT TTT GAA TTC TTG GTT CGC
    R   K   K   K   I   H   L   I   K   G   V   I   F   E   F   L   V   R

    TTA GGG TTT AAT AAT GGT AGA GAG AAA AGG TTG GAG GCA AAA GAA ATG AGC TTA
    L   G   F   N   N   G   R   E   K   R   L   E   A   K   E   M   S   L

    ACA ATT GAA CAT TTA ACA GGA GGT TAC GGC CAT ATT CCT GTC TTA AAA GAT ATT
    T   I   E   H   L   T   G   G   Y   G   H   I   P   V   L   K   D   I

    AAT TTT GAC GTC AAG TCT GGT GAA ATG GTT GGT TTA ATT GGC TTG AAT GGT GCG
    N   F   D   V   K   S   G   E   M   V   G   L   I   G   L   N   G   A

    GGG AAA AGT ACC ACC ATC AAA AAT ATT ATT GGG TTA CTT ACG CCG CAA AAA GGC
    G   K   S   T   T   I   K   N   I   I   G   L   L   T   P   Q   K   G

    AAA ATT ATG ATT GAC GGG GAA ACA TTG CAA CAA GCC CCT GAA GAA TAT CGT AAA
    K   I   M   I   D   G   E   T   L   Q   Q   A   P   E   E   Y   R   K

    AAA ATC GGG TAT ATT CCG GAA ACA CCT TCT TTA TAT GAA GAA TTA ACA TTG AAA
    K   I   G   Y   I   P   E   T   P   S   L   Y   E   E   L   T   L   K

    GAG CAT ATC GAA GTA ACG GCC TTG GCT TAT GAT ATT CCA TTA GAA GAA GCG TTC
    E   H   I   E   V   T   A   L   A   Y   D   I   P   L   E   E   A   F

    AAA CGA GCA GAA CCA TTA CTA AAA ACG TTC CGT TTA GAC AAT AAA TTG GAA TGG
    K   R   A   E   P   L   L   K   T   F   R   L   D   N   K   L   E   W

    TTT CCT GCT AAT TTT TCA AAA GGC ATG AAA CAA AAA GTC ATG GTA CTT TGT GCA
    F   P   A   N   F   S   K   G   M   K   Q   K   V   M   V   L   C   A

    TTC TTA ATT GAA CCG AGT TTA TAT ATT ATT GAT GAA CCT TTT CTA GGC TTA GAT
    F   L   I   E   P   S   L   Y   I   I   D   E   P   F   L   G   L   D

    CCT TTA GCA ATT CAT GCT TTA TTA GAA TTA ATG GAT ACG ATG CGT AAG CAA GGG
    P   L   A   I   H   A   L   L   E   L   M   D   T   M   R   K   Q   G

    GCA GCG ATT TTA ATG TCC ACG CAT ATT TTA GCA ACA GCT GAA AAA TAT TGT GAT
    A   A   I   L   M   S   T   H   I   L   A   T   A   E   K   Y   C   D

    CGT TTT GTG GTG CTA CAC GAA GGG AAA TTA CGG GCA AAC GGT ACA ATG GCT GAA
    R   F   V   V   L   H   E   G   K   L   R   A   N   G   T   M   A   E

    TTA CGT GCA GAA TTT AAT TTA CCA GAG TCT TCT TTA GAT GAT ATT TAT CTT GCC
    L   R   A   E   F   N   L   P   E   S   S   L   D   D   I   Y   L   A
```

Fig. 18A

```
TTG ACG AAG GAA GAA AAG GTG GGG TAA 3'
--- --- --- --- --- --- --- --- ---
 L   T   K   E   E   K   V   G   *
```

Fig. 18B

```
5'  ATG AAA AAA TTA AGT ATT CAT CTG AAA GAC GTA AGT ATT CAC TTT TCG GGA AAA
    M   K   K   L   S   I   H   L   K   D   V   S   I   H   F   S   G   K

    CCA ATT CTA GAA ATT GAT GAA TTA TTT GTA TAT GAA AAT GAA AAA ATT GGG ATT
    P   I   L   E   I   D   E   L   F   V   Y   E   N   E   K   I   G   I

    ATT GGT AAA AAC GGC GCA GGA AAA TCA ACA CTG CTC AAT TTA ATT ATG GGT AAG
    I   G   K   N   G   A   G   K   S   T   L   L   N   L   I   M   G   K

    ATT CAA TCA GAT AAA GGA AAA GTT CAA AGA TTG AAC GAC TTT CAT TAC TTG GCA
    I   Q   S   D   K   G   K   V   Q   R   L   N   D   F   H   Y   L   A

    CAA GTA GCT GAA GAA ATA ACG AAT GAA TCA GAG AAA ACT GAC AAA AAT TGT TTG
    Q   V   A   E   E   I   T   N   E   S   E   K   T   D   K   N   C   L

    CTA AAT CAG AAA AAT CAA AAA CTA AGT GGT GGC GAA AAA GTC CAA AAA CGC TTA
    L   N   Q   K   N   Q   K   L   S   G   G   E   K   V   Q   K   R   L

    GCA ACA TTA TTT TCA GAG TAT CCA ACC GGG GTT ATT TTA GAT GAA CCA ACC ACA
    A   T   L   F   S   E   Y   P   T   G   V   I   L   D   E   P   T   T

    CAT TTA GAT AAA GAA CAT CGT CAG TTG TTA GTG GCA GAT TTA ACG TAT TAT TAT
    H   L   D   K   E   H   R   Q   L   L   V   A   D   L   T   Y   Y   Y

    GGG ACC GTT TTG TTT GTT AGT CAT GAT CGC TTT TTC TTG AAT CAA TTA GCA GAG
    G   T   V   L   F   V   S   H   D   R   F   F   L   N   Q   L   A   E

    AAA ATT TGG GAA GTT TCG GAT GGA CAT GTC AAA GAA TAT TTA GGG AAC TAC GAT
    K   I   W   E   V   S   D   G   H   V   K   E   Y   L   G   N   Y   D

    GCG TAT TGT CGT CAA AAA GAA TTG GAA CAG CAA ACA CAA TAT AAT GTC TAT CAT
    A   Y   C   R   Q   K   E   L   E   Q   Q   T   Q   Y   N   V   Y   H

    CAG TAT CAA AAG GAA AAG AAA AAT TAC AGG AAT CTT ACG CAA AGA AAC AAG CAC
    Q   Y   Q   K   E   K   K   N   Y   R   N   L   T   Q   R   N   K   H

    AAG CGC AAA AAT CTA GTC ATG TTT CAA AAA AAC AWA AWC AAA AGC ARA TTA AAC
    K   R   K   N   L   V   M   F   Q   K   N   X   X   K   S   X   L   N

    CTA GTC GTT TAG 3'
    L   V   V   *
```

Fig. 19

```
5'  ATG ATT TTA TTA CAA GCA AAT CAA GTT GCC CGG CAT TTT GGC TCG GAA ACA TTG
    M   I   L   L   Q   A   N   Q   V   A   R   H   F   G   S   E   T   L

    TTT GAA AAC ATA CAT TTA GAA ATT GCA ACA AAA AGT CGG ATT GCC TTA GTT GGT
    F   E   N   I   H   L   E   I   A   T   K   S   R   I   A   L   V   G

    CGT AAT GGT GCT GGA AAA TCG ACT TTT TTA AAA ATC ATT GCA GGC ATT GAT GCT
    R   N   G   A   G   K   S   T   F   L   K   I   I   A   G   I   D   A

    CCC GAT AGC GGA ACC ATT GCC AAA AAT AAA ACT GCT ACG TTA GGT TAT TTA GCT
    P   D   S   G   T   I   A   K   N   K   T   A   T   L   G   Y   L   A

    CAA AAT ACC GGC TTA GAA TCA GAT AAA ACC GTT TGG GAA GAA ATG ACA AAA GCC
    Q   N   T   G   L   E   S   D   K   T   V   W   E   E   M   T   K   A

    TTT GCT GAC ATC CTA GAA ATG GAA CAG CGT ATG CGA GAA TTA GAA ACT AAA ATT
    F   A   D   I   L   E   M   E   Q   R   M   R   E   L   E   T   K   I

    AGT GAA ATG GAG CCA ACC ACT TCC GTT TAT GAA GGA ATT TTA AAA GAG TAC GAT
    S   E   M   E   P   T   T   S   V   Y   E   G   I   L   K   E   Y   D

    CAA TTG CAA CAT ACG TTT TCT GAA AAA AAT GGC TAC GGC TAT GAA AAT GAA ATT
    Q   L   Q   H   T   F   S   E   K   N   G   Y   G   Y   E   N   E   I

    CGC TCA GTC CTT CAC GGC TTT GGC TTT GAT GAA TCC TTT TAC ACG AAA GAT ATT
    R   S   V   L   H   G   F   G   F   D   E   S   F   Y   T   K   D   I

    CAA ACC TTA TCT GGT GGT CAA AAA ACC CGG CTT GCA TTA GCG AGA ATG CTT TTA
    Q   T   L   S   G   G   Q   K   T   R   L   A   L   A   R   M   L   L

    CAA AAA CCA GAC ATT TTA ATT CTG GAC GAG CCT ACA AAC CAC TTA GAT ATC GAG
    Q   K   P   D   I   L   I   L   D   E   P   T   N   H   L   D   I   E

    ACG CTT TCT TGG CTG GAA TCT TAT TTG CCA AGT TAT GCC GGC GCC CTA TTA ATT
    T   L   S   W   L   E   S   Y   L   P   S   Y   A   G   A   L   L   I

    GTT TCC CAC GAT CGT TAT TTT TTA GAT AAG GTA GTT AAT GAA GTT TAT GAA CTG
    V   S   H   D   R   Y   F   L   D   K   V   V   N   E   V   Y   E   L

    AGT CGC AAA AAA ATG ACT CAC TAC AAA GGA AAC TAT TCC AAA TAC TTA GAG TTA
    S   R   K   K   M   T   H   Y   K   G   N   Y   S   K   Y   L   E   L

    AAA GCA GAA CAA TTA GCC AGT GAA TGG AAA GCG TAT GAA AAG CAA CAA GAA GAA
    K   A   E   Q   L   A   S   E   W   K   A   Y   E   K   Q   Q   E   E

    ATC AAT AAG TTA GAA GAT TTC GTT GCC AAA AAT CTG GTT CGT GCA TCT ACA ACG
    I   N   K   L   E   D   F   V   A   K   N   L   V   R   A   S   T   T
```

Fig. 20A

```
AAA CGT GCA CAA AGT CGC CGA AAA GTA TTA GAA AAA ATG GAC CGT TTA GAC CGA
---------------------------------------------------------------
 K   R   A   Q   S   R   R   K   V   L   E   K   M   D   R   L   D   R

CCT CAA GGA GAT GAA AAA TCG GCG CAT TTT CTT TTC GAT AGT GAA AAA GTC TCG
---------------------------------------------------------------
 P   Q   G   D   E   K   S   A   H   F   L   F   D   S   E   K   V   S

GGA AAT GTT GTT TTA CAA GTC GAA GAT GCC GCC ATT GGT TAC GAC CAA GAA CAT
---------------------------------------------------------------
 G   N   V   V   L   Q   V   E   D   A   A   I   G   Y   D   Q   E   H

ATT TTA TCC GAA CCT ATT CAC TTG GAT ATT CGT CGC AAA GAA GCC ATT GCC TTA
---------------------------------------------------------------
 I   L   S   E   P   I   H   L   D   I   R   R   K   E   A   I   A   L

GTC GGA CCG AAC GGA ATT GGT AAA TCC ACT CTC TTG AAA TCA ATT ATT GAC CGC
---------------------------------------------------------------
 V   G   P   N   G   I   G   K   S   T   L   L   K   S   I   I   D   R

ATT CCT TTC ATT AAA GGA AGT AAA ACT TTT GGC ACC AAT GTT TCT GTA GGT TAC
---------------------------------------------------------------
 I   P   F   I   K   G   S   K   T   F   G   T   N   V   S   V   G   Y

TAT GAC CAA GAG CAA GCC AAT TTA CAT GGC AAT AAA ACG GTC TTA GCG GAA TTA
---------------------------------------------------------------
 Y   D   Q   E   Q   A   N   L   H   G   N   K   T   V   L   A   E   L

TGG GAT GAA CAC CCA ACC ACA CCT GAA AAA GAG ATT CGA AGT ATT TTA GGC GGC
---------------------------------------------------------------
 W   D   E   H   P   T   T   P   E   K   E   I   R   S   I   L   G   G

TTT CTC TTC AGT GGA GAC GAT GTT GAA AAA ACG ATT CCT TTA TTA AGC GGT GGC
---------------------------------------------------------------
 F   L   F   S   G   D   D   V   E   K   T   I   P   L   L   S   G   G

GAA AAA GCC CGT GTG GCA TTA GCA AAA CTA GCG ATG GAT CGT GAC AAT TTC TTG
---------------------------------------------------------------
 E   K   A   R   V   A   L   A   K   L   A   M   D   R   D   N   F   L

ATT CTC GAT GAG CCA ACC AAT CAC TTG GAT ATC GAT AAT AAA GAA GTT TTA GAA
---------------------------------------------------------------
 I   L   D   E   P   T   N   H   L   D   I   D   N   K   E   V   L   E

AAT GCG CTG ATT GAT TAT GAA GGA ACC ATC CTC TTC GTT TCC CAT GAC CGT TAC
---------------------------------------------------------------
 N   A   L   I   D   Y   E   G   T   I   L   F   V   S   H   D   R   Y

TTT ATC AAT CGA ATT GCA ACA AAA GTT GTT GAG CTT TCT GAA AAA GGC AGC AAA
---------------------------------------------------------------
 F   I   N   R   I   A   T   K   V   V   E   L   S   E   K   G   S   K

CTT TAT TTA GGC GAC TAT GAT TAT TAT TTA GAA AAG AAA CAA GAG GAA GAA GAA
---------------------------------------------------------------
 L   Y   L   G   D   Y   D   Y   Y   L   E   K   K   Q   E   E   E   E

ATC GCT GCC CTC TTA GCT AAT GAA GAA GCG GCG AAA AAA CCC GAA CCA GTT ACA
---------------------------------------------------------------
 I   A   A   L   L   A   N   E   E   A   A   K   K   P   E   P   V   T

GCC AAA AAT ACC TTT TAT CAA AAC AAG GAG CAA CAA AAA TTA CTC CGT ACT TTG
---------------------------------------------------------------
 A   K   N   T   F   Y   Q   N   K   E   Q   Q   K   L   L   R   T   L
```

Fig. 20B

```
CAA AGA AAA ATA ACA CAA GTC GAA GAA AAT CTT GCT CAG TTA GAT ACG ACA ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   R   K   I   T   Q   V   E   E   N   L   A   Q   L   D   T   T   I

GCA CAA TTA GAA GCA CAA ATG AGT CAA CCA GAC ATT CTA GAA AAT CAT GTT GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   Q   L   E   A   Q   M   S   Q   P   D   I   L   E   N   H   V   E

TTG CTG GCT TTA AAC CAA CAA TTA GAT GAA GTT CGT CAG CAA CAA GAT GAA CTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   A   L   N   Q   Q   L   D   E   V   R   Q   Q   Q   D   E   L

CTT GAG CAA TGG GAA AAT TTC AGT TTA GAA TTA GAA GAA ATG GAA AAT AAC AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   E   Q   W   E   N   F   S   L   E   L   E   E   M   E   N   N   N

TAA 3'
---
 *
```

Fig. 20C

```
5'  ATG AAT AAA CAA GAT GCT TGG ACA GCA GAT ACC AAT GCA AAA ATT ATT TTA CAA
    M   N   K   Q   D   A   W   T   A   D   T   N   A   K   I   I   L   Q

    AAA TTA GGC ATT CCA ACA TTG GAA AAG AAA ATA GGT GAG CTT TCT GGT GGG CAG
    K   L   G   I   P   T   L   E   K   K   I   G   E   L   S   G   G   Q

    TTA AAA CGA GTA AGC TTG GCG CAA GTC TTA ATT GAA GCG CCT GAT TTA CTG TTA
    L   K   R   V   S   L   A   Q   V   L   I   E   A   P   D   L   L   L

    TTG GAT GAA CCA ACG AAC CAT CTG GAT TAT GAA ACG ATT GAA TGG CTG GAA AAC
    L   D   E   P   T   N   H   L   D   Y   E   T   I   E   W   L   E   N

    TTT TTA AAT AAC TAT CGT GGC GCG GTT TTA ATG GTG ACC CAT GAC CGC TAC TTT
    F   L   N   N   Y   R   G   A   V   L   M   V   T   H   D   R   Y   F

    TTA GAT CGA GTA ACC AAT CGG ATT TTC GAA CTT TCT TTT GGA AAA CTG TAT GAA
    L   D   R   V   T   N   R   I   F   E   L   S   F   G   K   L   Y   E

    TAC AAA GGA AAC TAT GAA ACG TAT GTG ATG GAA AAA GCC GAA CGT GAA CGT GTC
    Y   K   G   N   Y   E   T   Y   V   M   E   K   A   E   R   E   R   V

    GCT GTA GAA CAA GAA GAA AAA AGA AAA CGC CTC TTC AAA CAA GAA CTT GCG TGG
    A   V   E   Q   E   E   K   R   K   R   L   F   K   Q   E   L   A   W

    ATG CGA GCA GGC GTT CAA GCA CGA GGA ACG AAA CAA CAA GCA AGA ATT GAT CGG
    M   R   A   G   V   Q   A   R   G   T   K   Q   Q   A   R   I   D   R

    TTT CAG GAT TTA AAA GAA AAC TTG CAT CAG GTA CAA ACC AAT GGC ACC TTA GAA
    F   Q   D   L   K   E   N   L   H   Q   V   Q   T   N   G   T   L   E

    GCA GAC TTT GCC ACG CAG CGT TTA GGG AAA AAA GTT CTG GAA ATC AAG GAA GGA
    A   D   F   A   T   Q   R   L   G   K   K   V   L   E   I   K   E   G

    AAT TAT GCC ATT GAT CAT AAA CAG CTT TTG AAA GAC TTT AAT TTA CTC ATT CAA
    N   Y   A   I   D   H   K   Q   L   L   K   D   F   N   L   L   I   Q

    GCA AAC GAT CGA ATT GGC ATT ACT GGT AAA AAC GGT GCA GGC AAA TCC ACA TTA
    A   N   D   R   I   G   I   T   G   K   N   G   A   G   K   S   T   L

    TTA AAT ATT TTA GCT GGA CGC ATT CCT TTA GAG AGT GGT CTA TAT AGT ATC GGT
    L   N   I   L   A   G   R   I   P   L   E   S   G   L   Y   S   I   G

    GAA ACC GTC CGA ATT GGC TAT TAT ACC CAG CAA AAC GAA GAA ATG GAT CCT AAT
    E   T   V   R   I   G   Y   Y   T   Q   Q   N   E   E   M   D   P   N

    CAA CGG ATG ATT GCT TAT TTG CAA GAA GCA GCA GAA GAA GTG AAA CGC AGT GAT
    Q   R   M   I   A   Y   L   Q   E   A   A   E   E   V   K   R   S   D
```

Fig. 21A

```
GGA AGC AGC GTT GGG GTC GCT GAA ATG TTA GAG CGC TTT TTA TTT CCT CGT TTT
---------------------------------------------------------------------
 G   S   S   V   G   V   A   E   M   L   E   R   F   L   F   P   R   F

ATG CAT GGC ACA CTG ATT GGC AAA CTG TCT GGT GGA GAA AAA CGC CGG TTA TAT
---------------------------------------------------------------------
 M   H   G   T   L   I   G   K   L   S   G   G   E   K   R   R   L   Y

CTC TTA AAA TTA TTA ATT AGT CAG CCT AAC GTG TTA TTA TTG GAT GAG CCA ACC
---------------------------------------------------------------------
 L   L   K   L   L   I   S   Q   P   N   V   L   L   L   D   E   P   T

AAT GAT TTG GAC ATT GAT ACG TTG ACC ATT TTA GAA GAT TAT ATC CAA ACG TTT
---------------------------------------------------------------------
 N   D   L   D   I   D   T   L   T   I   L   E   D   Y   I   Q   T   F

AAA GGA GCG GTA ATT GCG GTT TCT CAT GAC CGT TAT TTC TTG GAT AAA ACA ATG
---------------------------------------------------------------------
 K   G   A   V   I   A   V   S   H   D   R   Y   F   L   D   K   T   M

GAT AAA TTG TTA GTT TTT CAA GGA AAT GGT CAG ATC CTA TCA TTT TAT GGC ACG
---------------------------------------------------------------------
 D   K   L   L   V   F   Q   G   N   G   Q   I   L   S   F   Y   G   T

ATG AGT GAA TAT TTG GCG AAC CAA AAA GAA CAA GAA AAG GTC AAA GAA AAA CCG
---------------------------------------------------------------------
 M   S   E   Y   L   A   N   Q   K   E   Q   E   K   V   K   E   K   P

CTC AAT AAA CCA GTA AAA GAA GCA TCG ACA GAA AAG AAA GAA AAA ACA AAG CTG
---------------------------------------------------------------------
 L   N   K   P   V   K   E   A   S   T   E   K   K   E   K   T   K   L

ACT TAT ATG GAA CAA AAA GAG TGG GCA ACA ATT GAA GAA GAC ATC ACA CAA TTA
---------------------------------------------------------------------
 T   Y   M   E   Q   K   E   W   A   T   I   E   E   D   I   T   Q   L

GAA ACG AAA TTA GAA ACA TTA ACT GAA GAA ATG AAT CAT CAG GGT GAT GAT TTT
---------------------------------------------------------------------
 E   T   K   L   E   T   L   T   E   E   M   N   H   Q   G   D   D   F

GTT CGT TTA CAA GAA TTG CAA GAA GCA GTA ACA GCA ACG GAA CAA TTG TTA GAA
---------------------------------------------------------------------
 V   R   L   Q   E   L   Q   E   A   V   T   A   T   E   Q   L   L   E

GAA AAA ATG AAT CGT TGG GAA TAC TTG AGT GAA TAC GCG GAT TAA 3'
-------------------------------------------------------------
 E   K   M   N   R   W   E   Y   L   S   E   Y   A   D   *
```

Fig. 21B

```
5'  ATG AGT ATT TTA ACT ATT GAA CAT TTA ACG CAT CGA TTT GGC GAA AAG GTC TTG
    M   S   I   L   T   I   E   H   L   T   H   R   F   G   E   K   V   L

    TAT GAA GAG GCT TCA TTG CAA GTG AAT AAA GGG GAT CAT TTA GGC TTA ACT GGC
    Y   E   E   A   S   L   Q   V   N   K   G   D   H   L   G   L   T   G

    CAA AAT GGG GTC GGC AAA TCC ACC TTA ATT AAA ATT TTA ACG GGG GAA GTA TTG
    Q   N   G   V   G   K   S   T   L   I   K   I   L   T   G   E   V   L

    CCA GAC GAA GGA ACG ATT CAG TGG CAA AAA AAT TGC AAG ATT GGG TAC TTG GAT
    P   D   E   G   T   I   Q   W   Q   K   N   C   K   I   G   Y   L   D

    CAG CAC GTT TCT GTA GAA CAA TCA CTA ACG ATG GTT GAT TTT TTG AAA CAA GCC
    Q   H   V   S   V   E   Q   S   L   T   M   V   D   F   L   K   Q   A

    TTT CAA GAA CTT TTT GAT AAA GAA GCG AAA CTG ACA AAG CTT TAT GAA GAA TAC
    F   Q   E   L   F   D   K   E   A   K   L   T   K   L   Y   E   E   Y

    AGT CAA ACG GCT TCG GAA AAA CTT TTA GAA CAA GCA GGC AAG TTA CAA ACA GAT
    S   Q   T   A   S   E   K   L   L   E   Q   A   G   K   L   Q   T   D

    TTA GAT GAA AGT AAT TTT TAC CAA ATC GAC ACG ATT ATT CAG GAT TTA GCC AAT
    L   D   E   S   N   F   Y   Q   I   D   T   I   I   Q   D   L   A   N

    GGG TTA GGA CTA CAA GCA ATT GGT TTG GAT AAA AAG TTA GGG GAG CTA AGC GGT
    G   L   G   L   Q   A   I   G   L   D   K   K   L   G   E   L   S   G

    GGT CAA CGT TCA AAA GTG ATT TTA GCA AAA TTA TTG TTA GAG GCC CCT GAT GTG
    G   Q   R   S   K   V   I   L   A   K   L   L   L   E   A   P   D   V

    TTA CTT TTA GAT GAA CCC ACC AAC TAT TTA GAT GAT ACA CAT ATT CAA TGG CTT
    L   L   L   D   E   P   T   N   Y   L   D   D   T   H   I   Q   W   L

    GTT CGC TAT CTG AAT AAT TTT GAA GGA AGC TTT CTA TTA GTC TCC CAT GAT TAT
    V   R   Y   L   N   N   F   E   G   S   F   L   L   V   S   H   D   Y

    CAA TTT TTA AAT GAA GTG ACG AAT TGC ATC GCA GAC ATT GAG TTT GGC AAG TTA
    Q   F   L   N   E   V   T   N   C   I   A   D   I   E   F   G   K   L

    ACC AAA TAT ACT GGA AAT GTA GAA AAA TCT TTT GCA CAA AAA GAG CAG AAT AAA
    T   K   Y   T   G   N   V   E   K   S   F   A   Q   K   E   Q   N   K

    CAA ACC TAT TTG AAA CAG TAT CAG GCC CAA CAA GAA AAA ATT GAA AAA ATG GAA
    Q   T   Y   L   K   Q   Y   Q   A   Q   Q   E   K   I   E   K   M   E

    GCC TAT ATT CGT AAA TAC AAA GCT GGA AAT CGA GCA ACG ATG GCT AAA AGT CGA
    A   Y   I   R   K   Y   K   A   G   N   R   A   T   M   A   K   S   R
```

Fig. 22A

```
CAA AAA CAA TTG GAC CGG TTG GAA CGA TTG ACT CCG CCT GGT TCC TTG ACT AAG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Q    K    Q    L    D    R    L    E    R    L    T    P    P    G    S    L    T    K

CCA GCG ATT GAA TTT CCT TAT CAA GGG TTA GTT GCA ACG CAA GCA CTA ACG ACC
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 P    A    I    E    F    P    Y    Q    G    L    V    A    T    Q    A    L    T    T

CAG AAG TTA GTT GTC GGC TAT CGG GAA CCT TTG TTA GAA CCG TTA GAT TTA ATG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Q    K    L    V    V    G    Y    R    E    P    L    L    E    P    L    D    L    M

GTT CAT GTC GGT GAA AAA GTC GCA TTG AAA GGC TTT AAT GGG ATT GGC AAA TCA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 V    H    V    G    E    K    V    A    L    K    G    F    N    G    I    G    K    S

ACA TTA ATT AAA ACG TTG ACG AAA GTG ATT CCT TCA TTA GAT GGA GAA TTT CAT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 T    L    I    K    T    L    T    K    V    I    P    S    L    D    G    E    F    H

TAT CCG CTG AAT ACA AAA ATT GCT TAT TTT ACC CAA GAC TTA GCG TGG CCT AAT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Y    P    L    N    T    K    I    A    Y    F    T    Q    D    L    A    W    P    N

GAG CAG TTA ACG CCA CTA GAC TAT TTA TCA GAT CGT TTT CCA GAT ACA ACG ATT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 E    Q    L    T    P    L    D    Y    L    S    D    R    F    P    D    T    T    I

AAA GAG CGA AGA AGT CAT TTG GCT AGA GCA GGT TTA CCA GAT AAG TTA GCA ATG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 K    E    R    R    S    H    L    A    R    A    G    L    P    D    K    L    A    M

CAG TCG CTA GCC CTG TTA AGT GGT GGT GAG CAA ACA AAA GTA AAA CTA GCG GAA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 Q    S    L    A    L    L    S    G    G    E    Q    T    K    V    K    L    A    E

CTA ATG ATG CAA ACA AGT AAT TTG TTA TTT TTA GAT GAG CCA ACC AAT CAT ATT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 L    M    M    Q    T    S    N    L    L    F    L    D    E    P    T    N    H    I

GAT GAA GCG GCC AAA AAA ATT TAC AAG AAG CAA TTC ACG TTT ACC CAG GAA CGG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 D    E    A    A    K    K    I    Y    K    K    Q    F    T    F    T    Q    E    R

TTT TTC TGG WTT CCC ATG AAG CAG ATT TTT ATG AGG AAA TTG TGG ATC GAG TAA 3'
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 F    F    W    X    P    M    K    Q    I    F    M    R    K    L    W    I    E    *
```

Fig. 22B

```
5'  ATG ACT GAT GCT CTT GTG ATA CAA GAT TTA CGC AAA GTG TAC GCT TCA GGT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   T   D   A   L   V   I   Q   D   L   R   K   V   Y   A   S   G   V

    GAA GCG TTG CGG GGG ATT GAT TTA ACA GTC GAA GAA GGC GAT TTT TAT GCT CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   A   L   R   G   I   D   L   T   V   E   E   G   D   F   Y   A   L

    TTG GGA CCG AAT GGC GCA GGA AAA TCA ACG ACG ATT GGG ATT GTG ACC TCG TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   G   P   N   G   A   G   K   S   T   T   I   G   I   V   T   S   L

    GTT AAT AAG ACT TCG GGG AAA GTC AAA ATT TTT GGG TAT GAT TTA GAT ACG GAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   N   K   T   S   G   K   V   K   I   F   G   Y   D   L   D   T   E

    ATG GTG CGG GCG AAG CAA CAG ATT GGG TTA GTT CCG CAG GAG TTT AAC TTC AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   V   R   A   K   Q   Q   I   G   L   V   P   Q   E   F   N   F   N

    CCG TTT GAA ACG GTC CAG CAG ATT GTG GTC AAT CAA GCG GGC TAT TAT GGG GTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   F   E   T   V   Q   Q   I   V   V   N   Q   A   G   Y   Y   G   V

    TCT CGT AAA GAA GCC ATG AAA CGG AGT GAA AAG TAT TTA AAA CAG TCC AAT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   R   K   E   A   M   K   R   S   E   K   Y   L   K   Q   S   N   L

    TGG GAA AAA CGG AAC GAA CGG GCG CGA ATG CTT TCT GGC GGA ATG AAA CGT CGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     W   E   K   R   N   E   R   A   R   M   L   S   G   G   M   K   R   R

    TTG ATG ATT GCT CGC GCG CTG ATG CAT GAG CCG AAG TTA TTG ATT TTA GAT GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   M   I   A   R   A   L   M   H   E   P   K   L   L   I   L   D   E

    CCA ACT GCA GGT GTC GAT ATT GAA TTG CGC CGT GAA ATG TGG GCC TTT TTA CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   T   A   G   V   D   I   E   L   R   R   E   M   W   A   F   L   Q

    GAG TTA AAT GCG CAG GGA ACC ACG ATT ATT TTG ACG ACC CAT TAC TTA GAA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   L   N   A   Q   G   T   T   I   I   L   T   T   H   Y   L   E   E

    GCT GAG ATG TTG TGT CGC AAC ATT GGG ATT ATT CAG TCT GGT GAA TTG ATT GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   E   M   L   C   R   N   I   G   I   I   Q   S   G   E   L   I   E

    AAC ACC AGC ATG AAG CAC TTG TTA GCA AAA TTA CAA TTT GAA ACG TTT ATT TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   T   S   M   K   H   L   L   A   K   L   Q   F   E   T   F   I   F

    GAT TTA GCA CCT TAT ACG CAG GCG CCA GTT ATT GAA GGC TAT CAA AGC GTA TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   L   A   P   Y   T   Q   A   P   V   I   E   G   Y   Q   S   V   F

    GAG GAC GAG TTA ACG TTA GCG GTG GAA GTG GAA CGA AAC CAA GGA GTC AAT CAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   D   E   L   T   L   A   V   E   V   E   R   N   Q   G   V   N   H

    TTG TTT GAA CAG CTA AGT CAA CAA GGA ATC AAA GTA TTG TCA ATG CGG AAT AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   F   E   Q   L   S   Q   Q   G   I   K   V   L   S   M   R   N   K
```

Fig. 23A

```
TCC AAC CGT TTG GAA GAG CTC TTC TTG AAG ATT ACC GAG GAT ACG TAC CAA AGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   N   R   L   E   E   L   F   L   K   I   T   E   D   T   Y   Q   R

GAG GAT CAA CAT GTT TAG 3'
--- --- --- --- --- ---
 E   D   Q   H   V   *
```

Fig. 23B

```
5'  ATG GCA TTA ATT GAA TTA CGT CAT GTA AAA AAA GAG TTT TCC GGT AAA GCG GGT
    M   A   L   I   E   L   R   H   V   K   K   E   F   S   G   K   A   G

    AAA GTC ACG GCT TTA AAA GAT ATT GAT TTA ACA GTT GAA TCA GGT GAT ATT TAT
    K   V   T   A   L   K   D   I   D   L   T   V   E   S   G   D   I   Y

    GGC ATT ATT GGC TAT TCA GGT GCA GGA AAA AGT ACG TTG GTT CGC CTA TTG AAT
    G   I   I   G   Y   S   G   A   G   K   S   T   L   V   R   L   L   N

    GGA CTT GAA ACA CCT ACA GAA GGT GAA GTA GAA ATT CAA GGA CAA GAT ATT GCA
    G   L   E   T   P   T   E   G   E   V   E   I   Q   G   Q   D   I   A

    TTG TTG CCA AAT AAA GAA CTA CGA AAC TTC CGT AAA AAA ATT GGG ATG ATC TTT
    L   L   P   N   K   E   L   R   N   F   R   K   K   I   G   M   I   F

    CAA CAT TTT AAT TTA TTA TGG TCA CGG ACG GTC CTA GAA AAT ATC ATG CTC CCA
    Q   H   F   N   L   L   W   S   R   T   V   L   E   N   I   M   L   P

    TTG GAA ATT GCT GGC GTT CCA AAA CAA AAT CGG AAA AGC CGC GCA GAA GAG TTA
    L   E   I   A   G   V   P   K   Q   N   R   K   S   R   A   E   E   L

    ATC AAG TTG GTT GGT TTA GAA GGA CGG GAA ACC GCC TAC CCA AGT CAA CTC TCT
    I   K   L   V   G   L   E   G   R   E   T   A   Y   P   S   Q   L   S

    GGT GGG CAA AAG CAA CGA GTG GGG ATT GCC CGT GCC TTA GCT AAC AAT CCT GAC
    G   G   Q   K   Q   R   V   G   I   A   R   A   L   A   N   N   P   D

    ATT TTG CTT TGT GAT GAA GCA ACG AGT GCT TTG GAT CCG CAG ACA ACT GAT GAA
    I   L   L   C   D   E   A   T   S   A   L   D   P   Q   T   T   D   E

    GTG CTA GAA TTA CTA CTA AAA ATC AAC CAA GAA TTA AAT TTA ACG GTT GTA TTA
    V   L   E   L   L   L   K   I   N   Q   E   L   N   L   T   V   V   L

    ATT ACC CAC GAA ATG CAC GTG ATT CGT AAA ATT TGT AAT CGC GTA GCT GTG ATG
    I   T   H   E   M   H   V   I   R   K   I   C   N   R   V   A   V   M

    GAA TAT GGT GAA ATT GTT GAA GAA GGT AAA GTG ATC GAT ATT TTC AAA AAG CCT
    E   Y   G   E   I   V   E   E   G   K   V   I   D   I   F   K   K   P

    CAA ACA GAA ATT GCG AAA CGC TTT ATC CAA CAA GAA GCG GAT AAA AAC ATT GAA
    Q   T   E   I   A   K   R   F   I   Q   Q   E   A   D   K   N   I   E

    GAA ACG GAA CTG GTT GTT GAA GAA ATG TTG GAA CAA TAT CCG AAT GGA AAA ATT
    E   T   E   L   V   V   E   E   M   L   E   Q   Y   P   N   G   K   I

    GTT CGC CTG CTT TTC CAC GGT GAA CAA GCG AAA TTG CCA ATT ATC TCG CAT ATT
    V   R   L   L   F   H   G   E   Q   A   K   L   P   I   I   S   H   I
```

Fig. 24A

```
GTC CAA GAA TAT CAA GTA GAA GTT AGT ATT ATT CAA GGG AAC ATT CAG CAA ACA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   Q   E   Y   Q   V   E   V   S   I   I   Q   G   N   I   Q   Q   T

AAG CAG GGT GCA GTG GGT TCT CTT TAT ATT CAA TTA TTA GGC GAA GAG CAA AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   Q   G   A   V   G   S   L   Y   I   Q   L   L   G   E   E   Q   N

ATT CTA GCA GCC ATT GAA GGA TTA CGT AAA CTT CGT GTA GAA ACA GAG GTG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   L   A   A   I   E   G   L   R   K   L   R   V   E   T   E   V   I

GGA AAT GAA TAA 3'
--- --- --- ---
 G   N   E   *
```

Fig. 24B

```
5'  ATG GGT CCT AAT GAG CGA CTA GCA ACG CTG AAA CAA AAT CAC TTT GAC TAC GAA
    M   G   P   N   E   R   L   A   T   L   K   Q   N   H   F   D   Y   E

    GAC TAC ACT GTT TTA GAA ACT GTA ATT ATG GGA CAT AAA CGT CTT TAC GAA GTA
    D   Y   T   V   L   E   T   V   I   M   G   H   K   R   L   Y   E   V

    ATG AAA GAA AAA GAT GCT ATC TAT ATG AAA GAA GAT TTT TCA GAT GAA GAT GGG
    M   K   E   K   D   A   I   Y   M   K   E   D   F   S   D   E   D   G

    ATT CGT GCC GCA GAA CTA GAA GGC GAA TTT GCT GAA CTT GAC GGT TGG GAA GCA
    I   R   A   A   E   L   E   G   E   F   A   E   L   D   G   W   E   A

    GAA CCT GAA GCA GCT GTT TTA CTC CAA GGG CTA AAC ATT CCA GAA GAA TTA CAT
    E   P   E   A   A   V   L   L   Q   G   L   N   I   P   E   E   L   H

    GAT CAA AAA ATG AGC GAA TTA ACA GCT GGT CAA AAA GTC AAA GTA TTA TTA GCT
    D   Q   K   M   S   E   L   T   A   G   Q   K   V   K   V   L   L   A

    CAA TCT CTT TTT GGT AAA CCA GAT GTC TTA CTA TTA GAC GAG CCG ACA AAT GGT
    Q   S   L   F   G   K   P   D   V   L   L   L   D   E   P   T   N   G

    TTA GAC ACC CGC TCA ATC AAT TGG TTA GAA GAA TTT TTA ATC AAC TTT GAA AAT
    L   D   T   R   S   I   N   W   L   E   E   F   L   I   N   F   E   N

    ACC GTT ATC GTG GTT TCC CAT GAC CGT CAT TTC CTA AAC AAA GTC TGC ACT CAC
    T   V   I   V   V   S   H   D   R   H   F   L   N   K   V   C   T   H

    ATG GCA GAT TTA GAC TTT AGT AAA ATC AAA CTT TAC GTT GGT AAC TAT GAT TTC
    M   A   D   L   D   F   S   K   I   K   L   Y   V   G   N   Y   D   F

    TGG TTG GAA TCA AGC CAA CTA GCG ACA AAA TTG CAA GCA CAA TCT AAT GCC AAA
    W   L   E   S   S   Q   L   A   T   K   L   Q   A   Q   S   N   A   K

    AAA GAA GAA CAA ATC AAA GAA TTA CAA GAC TTT ATC GCT CGT TTT AGC GCC AAT
    K   E   E   Q   I   K   E   L   Q   D   F   I   A   R   F   S   A   N

    GCC TCA AAA TCA AAA CAA GCA ACG TCT CGT AAA AAA ATG TTA GAT AAA ATT ACC
    A   S   K   S   K   Q   A   T   S   R   K   K   M   L   D   K   I   T

    TTA GAT GAC ATT CAA CCT TCC TCT CGT CGT TAT CCA TTC GTT GGT TTT ACA CCA
    L   D   D   I   Q   P   S   S   R   R   Y   P   F   V   G   F   T   P

    GAA CGT GAA ATC GGC AAT GAC TTA TTA CAA GTT GAA AAT GTC TCT GTA ACG ATT
    E   R   E   I   G   N   D   L   L   Q   V   E   N   V   S   V   T   I

    GAT GGC AAA AAA ATC TTA GAT AAT ATC TCG TTC AAC TTA ACA AAA GAT GAT AAA
    D   G   K   K   I   L   D   N   I   S   F   N   L   T   K   D   D   K
```

Fig. 25A

```
GTG GCA TTC ATC GCT GAC TCA GAT ATT ACA ACG ACT ACG TTA TTC AAA GTG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   A   F   I   A   D   S   D   I   T   T   T   T   L   F   K   V   I

ATG GGC GAA ATT ACT CCC GAT ACA GGT TCT GTT CGT TGG GGC GTT ACA ACT AGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 M   G   E   I   T   P   D   T   G   S   V   R   W   G   V   T   T   S

CAA GCT TAT TTA CCA AAA GAC AAC AGC AAA GAC TTT GAA GAG CCA TTA ACC ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   A   Y   L   P   K   D   N   S   K   D   F   E   E   P   L   T   I

TTA GAT TGG TTA CGT CAA TTT GCT GGT AAA GAA GAA GAT GAC AAT ACG TTC TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   D   W   L   R   Q   F   A   G   K   E   E   D   D   N   T   F   L

CGT AGT TTC TTA GGT CGG ATG TTA TTC TCT GGT GAA GAG GTA CTA AAA CCA GTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   S   F   L   G   R   M   L   F   S   G   E   E   V   L   K   P   V

AAT GTT CTT TCC GGA GGC GAA AAA GTG CGT GTC ATG CTT TCA AAA TTA ATG CTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   V   L   S   G   G   E   K   V   R   V   M   L   S   K   L   M   L

TCC AAA GCC AAT GTC TTA GTT TTA GAT GAT CCA ACG AAC CAC TTA GAC TTA GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   K   A   N   V   L   V   L   D   D   P   T   N   H   L   D   L   E

TCA ATC ACT GCA TTA AAT GAT GGG TTG ATG GCT TTC ACT GGT TCA ATT CTT TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   I   T   A   L   N   D   G   L   M   A   F   T   G   S   I   L   F

GCT TCA CAT GAC CAC CAA TTT ATC CAA ACA TTA GCG AAC CGA ATT ATT GCT GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   S   H   D   H   Q   F   I   Q   T   L   A   N   R   I   I   A   V

TCT GAT AAA GGG GTC ATT GAT CGG GCA GAA ACA ACC TAT GAT GAA TTC TTG GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   D   K   G   V   I   D   R   A   E   T   T   Y   D   E   F   L   E

AAC CCA GAA ATT CAA AAA CAA ATG GAC GTT TTA TTC AGT TCA GAT TAT TAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   P   E   I   Q   K   Q   M   D   V   L   F   S   S   D   Y   *
```

Fig. 25B

```
5'  ATG TCG AAA ATT GAA CTA AAA CAA CTA TCT TTT GCC TAT GAT AAT CAA GAA GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   S   K   I   E   L   K   Q   L   S   F   A   Y   D   N   Q   E   A

    TTG CTT TTT GAT CAG GCA AAT ATC ACG ATG GAT ACC AAT TGG AAA TTA GGA TTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   L   F   D   Q   A   N   I   T   M   D   T   N   W   K   L   G   L

    ATT GGC CGC AAT GGC CGT GGG AAA ACA ACC TTA TTA AGA TTG TTA CAA AAG CAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   G   R   N   G   R   G   K   T   T   L   L   R   L   L   Q   K   Q

    TTG GAT TAC CAA GGA GAG ATT CTT CAT CAA GTC GAT TTC GTC TAT TTT CCA CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   D   Y   Q   G   E   I   L   H   Q   V   D   F   V   Y   F   P   Q

    ACA GTT GCA GAA GAA CAA CAG CTC ACT TAT TAT GTC TTA CAA GAG GTG ACT TCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   V   A   E   E   Q   Q   L   T   Y   Y   V   L   Q   E   V   T   S

    TTT GAA CAG TGG AAA TTA GAA CGA GAA TTA ACG CTT TTA AAC GTT GAT CCT GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   E   Q   W   K   L   E   R   E   L   T   L   L   N   V   D   P   E

    GTT TTA TGG CGT CCC TTT TCT TCT TTA TCA GGC GGC GAA AAG ACG AAA GTT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   L   W   R   P   F   S   S   L   S   G   G   E   K   T   K   V   L

    TTA GGT CTT CTT TTT ATT GAA GAA AAT GCC TTT CCT TTA ATT GAC GAG CCA ACA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   G   L   L   F   I   E   E   N   A   F   P   L   I   D   E   P   T

    AAT CAC TTA GAT CTA GCT GGC AGA CAA CAA GTG GCT GAA TAT TTG AAG AAA AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   H   L   D   L   A   G   R   Q   Q   V   A   E   Y   L   K   K   K

    AAA CAC GGG TTT ATT TTA GTC AGC CAC GAT CGG GCA TTT GTT GAT GAA GTG GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   H   G   F   I   L   V   S   H   D   R   A   F   V   D   E   V   V

    GAT CAT ATT TTG GCG ATT GAA AAA AGT CAA TTG ACG CTG TAT CAA GGG AAT TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   H   I   L   A   I   E   K   S   Q   L   T   L   Y   Q   G   N   F

    TCT ATT TAT GAA GAG CAA AAA AAA TTA AGA GAT GCT TTT GAA CTA GCA GAA AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   I   Y   E   E   Q   K   K   L   R   D   A   F   E   L   A   E   N

    GAA AAA ATC AAA AAA GAA GTC AAT CGC TTG AAA GAA ACC GCT CGT AAA AAA GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   K   I   K   K   E   V   N   R   L   K   E   T   A   R   K   K   A

    GAA TGG TCG ATG AAC CGT GAA GGT GAT AAG TAC GGC AAC GCT AAG GAA AAA GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   W   S   M   N   R   E   G   D   K   Y   G   N   A   K   E   K   G

    AGC GGG GCG ATT TTT GAT ACA GGA GCC ATT GGT GCC CGG GCA GCG CGC GTA ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   G   A   I   F   D   T   G   A   I   G   A   R   A   A   R   V   M

    AAG CGC TCG AAA CAT ATT CAA CAA CGC GCC GAA ACA CAA TTA GCA GAA AAA GAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   R   S   K   H   I   Q   Q   R   A   E   T   Q   L   A   E   K   E
```

Fig. 26A

```
AAA CTA TTA AAA GAT CTT GAG TAT ATT GAT TCT TTG TCA ATG GAT TAT CAG CCA
 K   L   L   K   D   L   E   Y   I   D   S   L   S   M   D   Y   Q   P

ACG CAT CAC AAA ACA TTA TTG ACG GTG GAA GAG CTT CGT CTA GGC TAC GAG AAA
 T   H   H   K   T   L   L   T   V   E   E   L   R   L   G   Y   E   K

AAT TGG CTG TTT GCG CCA ATT TCT TTT TCA ATA AAC GCG GGA GAA ATT GTC GGA
 N   W   L   F   A   P   I   S   F   S   I   N   A   G   E   I   V   G

ATA ACA GGA AAA AAT GGC TCA GGA AAA TCG AGC TTG ATT CAG TAT TTA TTG GAT
 I   T   G   K   N   G   S   G   K   S   S   L   I   Q   Y   L   L   D

AAT TTT TCT GGG GAT TCA GAA GGC GAA GCC ACT TTG GCT CAC CAA TTA ACC ATT
 N   F   S   G   D   S   E   G   E   A   T   L   A   H   Q   L   T   I

TCT TAT GTG CGC CAA GAT TAT GAA GAC AAT CAA GGA ACT TTA TCC GAA TTT GCA
 S   Y   V   R   Q   D   Y   E   D   N   Q   G   T   L   S   E   F   A

GAG AAA AAT CAG TTA GAT TAC ACC CAA TTT TTA AAT AAC TTA CGA AAA CTT GGG
 E   K   N   Q   L   D   Y   T   Q   F   L   N   N   L   R   K   L   G

ATG GAG CGT GCC GTT TTC ACT AAT CGA ATT GAA CAA ATG AGT ATG GGG CAA CGG
 M   E   R   A   V   F   T   N   R   I   E   Q   M   S   M   G   Q   R

AAA AAA GTC GAA GTA GCC AAA TCA TTG TCT CAA TCA GCA GAG CTT TAT ATT TGG
 K   K   V   E   V   A   K   S   L   S   Q   S   A   E   L   Y   I   W

GAT GAA CCC CTT AAT TAC TTG GAT GTG TTT AAT CAT CAA CAA TTA GAA GCG CTA
 D   E   P   L   N   Y   L   D   V   F   N   H   Q   Q   L   E   A   L

ATC TTA TCT GTG AAG CCT GCA ATG CTA GTG ATT GAG CAT GAT GCG CAT TTC ATG
 I   L   S   V   K   P   A   M   L   V   I   E   H   D   A   H   F   M

AAG AAA ATA ACG GAT AAA AAA ATT GTC TTG AAA TCA TAA 3'
 K   K   I   T   D   K   K   I   V   L   K   S   *
```

Fig. 26B

```
5'  ATG CCG TTT ATT CCA GTA TAT GTT GAA CAA TTA GGG ACG CCT AAG AGT CAA GTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   P   F   I   P   V   Y   V   E   Q   L   G   T   P   K   S   Q   V

    GAG TTG TTT TCT GGA TTG GCA ATT TCT GTC ACG GCG TTT GCT GCT GCG ATT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   L   F   S   G   L   A   I   S   V   T   A   F   A   A   A   I   V

    GCA CCG ATT TGG GGA AAT TTA GCT GAT CGA AAA GGT CGG AAA ATT ATG ATG ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   P   I   W   G   N   L   A   D   R   K   G   R   K   I   M   M   I

    CGG GCT GCA GCA GGA ATG ACG ATC ACA ATG GGG GCG TTA GCA TTT GTC CCT AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   A   A   A   G   M   T   I   T   M   G   A   L   A   F   V   P   N

    GTA TAC TGG CTG TTG ATT ATG CGT TTT ATC AAC GGG ATT TTA TCC GGC TAT ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   Y   W   L   L   I   M   R   F   I   N   G   I   L   S   G   Y   I

    CCT AAC GCA ACA GCG ATG ATT GCG TCG CAG GCG CCA AAA GAG AAA AAT GGC TGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   N   A   T   A   M   I   A   S   Q   A   P   K   E   K   N   G   W

    GCT CTA GGG ACA TTA TCA ACA GGT GCA ATT GCT GGG ACG TTA ATT GCG CCA TCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   L   G   T   L   S   T   G   A   I   A   G   T   L   I   A   P   S

    ATT GGC GGT GCG TTG GCA CAG TGG TTT GGA ATG GAA AAT GTC TTT ATT ATT ACC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   G   G   A   L   A   Q   W   F   G   M   E   N   V   F   I   I   T

    GGC GTT ATT TTA TTC ATT ACA ACG TTG TTG ACT ATC TTT TTA GTT AAA GAA GAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   V   I   L   F   I   T   T   L   L   T   I   F   L   V   K   E   D

    TTT CAA CCA GTT GAG AAA AAG GAT TTA TTA ACG ACG AAA GAA ATT TTT AGT AAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   Q   P   V   E   K   K   D   L   L   T   T   K   E   I   F   S   K

    ATG GAT CAT GTT TCA GTA TTG ATT GGT TTG TTT GTG ACT ACG TTA ATT TTA CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   D   H   V   S   V   L   I   G   L   F   V   T   T   L   I   L   Q

    TTA GGA ATT ACA AGT ATC AGC CCA ATC TTA ACG CTA TAT ATT CGA TCT TTA AGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   G   I   T   S   I   S   P   I   L   T   L   Y   I   R   S   L   S

    GGC GAT ACA GAG AAT GTA TTA TTT GTT AGT GGT TTG ATT GTT TCG ATT GCC GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   D   T   E   N   V   L   F   V   S   G   L   I   V   S   I   A   G

    GTT TCA GCA ATT ATT TCT TCT CCA ACT TTG GGA AAA ATT GGG GAC CGC ATT GGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   S   A   I   I   S   S   P   T   L   G   K   I   G   D   R   I   G

    AAC CAA AAA GTT TTA TTA GGC GGA TTA ATT CTT TCT TTT ATT TGT TAC ATT CCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     N   Q   K   V   L   L   G   G   L   I   L   S   F   I   C   Y   I   P

    ATG GCG TTT GTG CAA ACA CCT TTT CAG TTA GGT GTT TTA CGA TTC CTT TTA GGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   A   F   V   Q   T   P   F   Q   L   G   V   L   R   F   L   L   G
```

Fig. 27A

```
TTT TCA ACA GGT GCT TTA ATG CCA TCA ATT AAT ACG TTA ATT AGT AAA ATA ACG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   S   T   G   A   L   M   P   S   I   N   T   L   I   S   K   I   T

CCA ACA GAG GGC GTA AGT AGA GTT TAT AGT TAC AAT CAG ATG TGT AGT AAT TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   T   E   G   V   S   R   V   Y   S   Y   N   Q   M   C   S   N   F

GGT CAA GTT TTA GGA CCG ATG GTA GGG TCG ACA GTT GCC CAT GGT TTT GGC TAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   Q   V   L   G   P   M   V   G   S   T   V   A   H   G   F   G   Y

TCT TCT GTT TTT CTG GTA ACC GCT TGT TTC GTT TTA GGA AAC ATT GGG CTG TCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   S   V   F   L   V   T   A   C   F   V   L   G   N   I   G   L   S

TTC TTT AAT TTC CGA AAA GTT TTA AAT AAA AAG CTC TAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- ---
 F   F   N   F   R   K   V   L   N   K   K   L   *
```

Fig. 27B

```
5'  ATG GCT GTT TTT TTA TGG AGG AGG CAA ATT ATG ACG AAA AAA AAT AGT ATG ATG
    M   A   V   F   L   W   R   R   Q   I   M   T   K   K   N   S   M   M

    TAC TTA GCA ATT TCT AAC TTA TTT CTT GTT TTT CTA GGC GTA GGC CTA GTC ATT
    Y   L   A   I   S   N   L   F   L   V   F   L   G   V   G   L   V   I

    CCC GTA ATT CCC CAA TTA AAA GAA GAA ATG CAT TTT TCT GGT ACC ACA ATG GGA
    P   V   I   P   Q   L   K   E   E   M   H   F   S   G   T   T   M   G

    ATG ATG ATT TCT ATT TTT GCG ATT GCC CAA TTA ATC ACA TCG CCA ATC GCA GGT
    M   M   I   S   I   F   A   I   A   Q   L   I   T   S   P   I   A   G

    GTC CTT TCG GAT AAA ATT GGT CGG AAA AAA ATG ATT GCA ACG GGC ATG TTG GTG
    V   L   S   D   K   I   G   R   K   K   M   I   A   T   G   M   L   V

    TTT TCA ATT TCT GAG TTA TTA TTT GGT TTA GCC CAA GCG AAA AGC GGT TTT TAT
    F   S   I   S   E   L   L   F   G   L   A   Q   A   K   S   G   F   Y

    ATT TCT CGT GGT TTA GGC GGG ATT GCC GCC GCT TTA TTA ATG CCG TCA GTG ACA
    I   S   R   G   L   G   G   I   A   A   A   L   L   M   P   S   V   T

    GCC TTT GTG GCA GAT ATG ACC ACG ATT TCT GAA CGT CCG AAA GCG ATG GGG CTT
    A   F   V   A   D   M   T   T   I   S   E   R   P   K   A   M   G   L

    GTG TCA GCC GCA ATT AGT GGT GGT TTT ATT ATC GGA CCA GGA GTT GGT GGT TTT
    V   S   A   A   I   S   G   G   F   I   I   G   P   G   V   G   G   F

    ATT GCT TAT TTA GGT ATT CGC GCC CCG TTT TTT GCA GCC GCA TTT TTA GCG TTT
    I   A   Y   L   G   I   R   A   P   F   F   A   A   A   F   L   A   F

    ATT GGT TTT ATT TTG ACA TTA ACT GTT TTG AAG GAG CCA GAG AAA CGA ATT TTA
    I   G   F   I   L   T   L   T   V   L   K   E   P   E   K   R   I   L

    GCC GCT GTT GAA GCG AAA AAA GGT TCA TTT ATG GAT ATT TTA AGA AAT CCA ATG
    A   A   V   E   A   K   K   G   S   F   M   D   I   L   R   N   P   M

    TTT ACC TCA TTA TTT GTG ATT ATC TTA ATT TCC TCT TTT GGC CTG CAA GCG TTC
    F   T   S   L   F   V   I   I   L   I   S   S   F   G   L   Q   A   F

    GAA TCT ATT TAT AGT ATT ATG GCG ACT ATT AAT TTT GGC TTT ACC ACA AGT GAA
    E   S   I   Y   S   I   M   A   T   I   N   F   G   F   T   T   S   E

    ATA GCA ATC GTG ATT ACG GTT AGT GGT ATT TTA GCG TTG ATT TGT CAG TTG TTT
    I   A   I   V   I   T   V   S   G   I   L   A   L   I   C   Q   L   F

    TTC TTT GAT GCA ATC GTC CAA AAA ATA GGT GAA ATG GGT TTA ATC CAA TTA ACC
    F   F   D   A   I   V   Q   K   I   G   E   M   G   L   I   Q   L   T
```

Fig. 28A

```
TTT TTT GCA AGT GCC ATT TTT ATT GCC GTG ATT GCC TTT ACA AAA AAT AAT TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   F   A   S   A   I   F   I   A   V   I   A   F   T   K   N   N   L

GTT GTT GTA TTT TCA ACG TTT ATT GTC TTT TTA GCG TTT GAC TTG TTT AGA CCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   V   V   F   S   T   F   I   V   F   L   A   F   D   L   F   R   P

GCA GTA ACT ACT TAT TTA TCC AAA CAT GCT GGA GAT CAA CAA GGA ACC ATC AAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   V   T   T   Y   L   S   K   H   A   G   D   Q   Q   G   T   I   N

GGA CTA AAT TCG ACA TTT ACA AGT TTT GGT AAT ATT TTA GGA CCA ATG GCA GCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   L   N   S   T   F   T   S   F   G   N   I   L   G   P   M   A   A

GGA GCT TTA TTT GAT ATC AAT CAC TTT TTC CCT TAT TAT GTT TCA GCA GTA ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   A   L   F   D   I   N   H   F   F   P   Y   Y   V   S   A   V   I

CTG TTA GGA ACG GGC TTT TTA TCG TTA TTT TTA AAT CGA AAT AAG ATG.TAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   G   T   G   F   L   S   L   F   L   N   R   N   K   M   *
```

Fig. 28B

```
5'  ATG ATC GAT AGA AAA AAA GTT ATT TTG TAT ACT TGT TGC ATG AGT TTG TTT GTA
    M   I   D   R   K   K   V   I   L   Y   T   C   C   M   S   L   F   V

    GTG ACC ATG GAT GTT ACG GTT GTC AAT GTG GCG TTA CCA TCC ATT CAA AGT GAT
    V   T   M   D   V   T   V   V   N   V   A   L   P   S   I   Q   S   D

    TTT CAC ACG AAT CTG TCT ACA TTA CAG TGG GTA ACA GAT GGC TAC ACT TTA ATG
    F   H   T   N   L   S   T   L   Q   W   V   T   D   G   Y   T   L   M

    GTA GCA TCC TTA TTA TTG TTG TCT GGG TCT ACA GCA GAT CGA ATT GGC CGT AAA
    V   A   S   L   L   L   L   S   G   S   T   A   D   R   I   G   R   K

    CGA GTC CTT CAA TTG GGC TTA GCC TGT TTT GGT TTA GCG TCT TTC CTA TGT GGG
    R   V   L   Q   L   G   L   A   C   F   G   L   A   S   F   L   C   G

    ATT TCG CAA ACC CCA GGG CAA TTG ATC GCG TTT CGC ATG TTG CAA GGG ATT GGT
    I   S   Q   T   P   G   Q   L   I   A   F   R   M   L   Q   G   I   G

    GGT TCT ATG TTA AAT CCG GTA GCA ATG TCT ATT ATT ACA CAG GTA TTT ACC GAA
    G   S   M   L   N   P   V   A   M   S   I   I   T   Q   V   F   T   E

    AAG TTA GAA AGA GCG AAA GCA ATT GGC TTG TGG GGC TCT GTT ACA GGG ATT TCC
    K   L   E   R   A   K   A   I   G   L   W   G   S   V   T   G   I   S

    TTA GGC ATG GGC CCG ATT ATT GGT GGA CTG ATT GTT TCT TAT TTT AGC TGG CGG
    L   G   M   G   P   I   I   G   G   L   I   V   S   Y   F   S   W   R

    TAT GTC TTT TTT GTA AAT GTA CCG ATC ATT GCT GCG GCA ATC ATC CTT ACA CAA
    Y   V   F   F   V   N   V   P   I   I   A   A   A   I   I   L   T   Q

    AAG TTT GTA CCT GAG TCA AAA GTA GAG AAG ACG GTG AAA AAT GAT TTT GTT GGT
    K   F   V   P   E   S   K   V   E   K   T   V   K   N   D   F   V   G

    CAA GCA TTG ATG ATT CTT TTT CTA TTT AGT TCT ATC TAT TCC ATT ATC GGA CTA
    Q   A   L   M   I   L   F   L   F   S   S   I   Y   S   I   I   G   L

    CCT AGA AAA GGG CTT TTC GCG CCA GAT ATT TTA AGT ACT GGG ATA ATC GGC TGC
    P   R   K   G   L   F   A   P   D   I   L   S   T   G   I   I   G   C

    TTA GCC ATT GTT ATT TTC TTT ATT TAT GAA TAT AAC ATT GAC AAT CCG TTA ATC
    L   A   I   V   I   F   F   I   Y   E   Y   N   I   D   N   P   L   I

    AAC CCG CGT TTC TTT TTA TCT ATT CCA TTT ACA TCG GCT TCT TTT TTA GCT ATT
    N   P   R   F   F   L   S   I   P   F   T   S   A   S   F   L   A   I

    TTT GGC TTT ATC ATA TAT AAC GGC TAT TTA TTT TTA AAC ACG CTA TAT TTG CAA
    F   G   F   I   I   Y   N   G   Y   L   F   L   N   T   L   Y   L   Q
```

Fig. 29A

```
GAG ATG AGA GGC TTC AGC CCG TTG GAA GCT GGC TTA TCA ACC ATT CCT TTG GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   M   R   G   F   S   P   L   E   A   G   L   S   T   I   P   L   A

TTT GTT AGT TTT CTT GTT GCA CCG AGA GCC GGC GAA ATG GTA GGG AGA ATA GGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   V   S   F   L   V   A   P   R   A   G   E   M   V   G   R   I   G

ACG AAA CGT CCT ATT ATG CTT TGT GGT ATT TCA ATG TTG GCT GTT AGC TTT TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   K   R   P   I   M   L   C   G   I   S   M   L   A   V   S   F   L

CAA TTA TTT GTA ACT AAA ACA ACG CCT ATG ATT ATT TTA TTT ATT ATT TAT ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   L   F   V   T   K   T   T   P   M   I   I   L   F   I   I   Y   I

TTT TTA GGC ATT GGG TTT GGG ATG TTA AAT TCA CCG ATT ACC ATT ACA GCG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   L   G   I   G   F   G   M   L   N   S   P   I   T   I   T   A   I

GAA GGA ATG CCA CTT TCT CAG TCA GGA ACT GCA GCA GCC ATT GCG GTG ACA TGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   G   M   P   L   S   Q   S   G   T   A   A   A   I   A   V   T   C

AAG CAA ATT GGC AAT TCT TTA GGG GTG GCG CTA CCA AGT CTT TTA ATT ACA AAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   Q   I   G   N   S   L   G   V   A   L   P   S   L   L   I   T   K

CCT ATT ACT AGT TCG CTT ACT CGA ACA CCT TTT ACA AAC GTA TGG CTT TTA TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   I   T   S   S   L   T   R   T   P   F   T   N   V   W   L   L   F

GGA TGT TGC GCC ATT GCG ATT ATC TTT TTA AGT TAT T 3'
--- --- --- --- --- --- --- --- --- --- --- --- -
 G   C   C   A   I   A   I   I   F   L   S   Y
```

Fig. 29B

```
5'  ATG GCA AAA GAA ACA AAT GTT AAG TTA GTC ACG GTG AGT GTT TTT GTG GCA ACA
    M   A   K   E   T   N   V   K   L   V   T   V   S   V   F   V   A   T

    TTT ATG ACA GCC ATT GAA GGG ACC ATT GTG TCT ACT GCG ATG CCA ACG ATT GTC
    F   M   T   A   I   E   G   T   I   V   S   T   A   M   P   T   I   V

    GGC TCG TTA CAT GGC ATG GAA ATT ATG AAC TGG GTA TTT TCA ATT TAT TTA TTA
    G   S   L   H   G   M   E   I   M   N   W   V   F   S   I   Y   L   L

    ACG AAT GCG ATG TTA ACA CCG ATT TAT GGG AAA CTT GCG GAT AAA ATT GGT CGT
    T   N   A   M   L   T   P   I   Y   G   K   L   A   D   K   I   G   R

    AAA CCT GTC TTC ATG ATT GGC ATT ATC ATT TTT ATT TTG GGC TCC TCG TTG TGT
    K   P   V   F   M   I   G   I   I   I   F   I   L   G   S   S   L   C

    GGC TTT GCT CAA GAT ATG TTG ACT TTA ATT ATT GCC CGC GCA ATT CAA GGT GTG
    G   F   A   Q   D   M   L   T   L   I   I   A   R   A   I   Q   G   V

    GGG GCA GGC GCA ATT TTA CCA GTT GCG TTA ACG ATT ATT GCC GAT ATG TAT ACA
    G   A   G   A   I   L   P   V   A   L   T   I   I   A   D   M   Y   T

    TTG GAC AAG CGA GCG AAA ATT TTA GGT TTA AAC AGT GCC GCC TGG GGA ATT GCT
    L   D   K   R   A   K   I   L   G   L   N   S   A   A   W   G   I   A

    AGT ATT TTT GGT CCG TTA GCA GGT GGT TTT ATT GTA GAT ACA GTC GGT TGG CAT
    S   I   F   G   P   L   A   G   G   F   I   V   D   T   V   G   W   H

    TGG ATT TTC TTC ATT AAT GTT CCT ATT GGA CTT GTT TTA TTG GGC TTG ATT AGT
    W   I   F   F   I   N   V   P   I   G   L   V   L   L   G   L   I   S

    ATT TTC TTA GTT GAA CCA AAG CGG GAA CGG ACC AAG ATG CCA ATG GAT ATT TTG
    I   F   L   V   E   P   K   R   E   R   T   K   M   P   M   D   I   L

    GGC AGC GTT ACT TTG ATG GTA GTG CTG CTA ACG TTA TTG CTA GGT TTT CAA ATG
    G   S   V   T   L   M   V   V   L   L   T   L   L   L   G   F   Q   M

    ATT AGC GAT AAT GGT TTT ACA TTA GTA ACA TTT GGT TGT TTA AGT TTG AGT GTG
    I   S   D   N   G   F   T   L   V   T   F   G   C   L   S   L   S   V

    CTC TTT TTT GTA GCA TTT GTG ATG ATA GAA AAA CGC GCG CAA GAC CCA GTG ATT
    L   F   F   V   A   F   V   M   I   E   K   R   A   Q   D   P   V   I

    GAT TTG CAT TTA TTT AAT CAA CCA ACG TTT GTT TTA GTA AAT CTT ATT GCA GCG
    D   L   H   L   F   N   Q   P   T   F   V   L   V   N   L   I   A   A

    CTT ATT AGC GGT TTC TTA ATG GGG ATT GAT GTC TAC ATT CCG ATG TGG ATG CAA
    L   I   S   G   F   L   M   G   I   D   V   Y   I   P   M   W   M   Q
```

Fig. 30A

```
GGT GTC TTA GGA AAA AGT GCA GGA ATT GGT GGC TTA GTT TTA GCG CCT ATG TCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   V   L   G   K   S   A   G   I   G   G   L   V   L   A   P   M   S

TTA CTT TGG ATG GCT GGA TCA TTT ATA GCA AGT AGT TTT ATG GAA AAA TAT GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   W   M   A   G   S   F   I   A   S   S   F   M   E   K   Y   A

ATG AAA AAA GTC TTA ACG ATT GGG TTA TCG ATT CTA TTA GTC GGG GCC ATC TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 M   K   K   V   L   T   I   G   L   S   I   L   L   V   G   A   I   F

TTA GTG GTA ATG CCA ATG GCC GTT CCG TTT TGG CTT TTC TTT GTA GTG TCT TCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   V   V   M   P   M   A   V   P   F   W   L   F   F   V   V   S   S

GTC TTA GGA GTT GGT TTT GGG ATT ACT ATC ACA ACG ACT ACG GTG ACA GCA CAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   L   G   V   G   F   G   I   T   I   T   T   T   V   T   A   Q

AGT ACA GTG GAG CCT GAA AAA ATG GGG GTT GCA ACA TCG TTT AAT ACG TTG GTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   T   V   E   P   E   K   M   G   V   A   T   S   F   N   T   L   V

CGT ACA ATT GGG CAG ACT GTG ATG GTG TCA ATT TTT GGT GTG ATT TTA AAT GCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   T   I   G   Q   T   V   M   V   S   I   F   G   V   I   L   N   A

GGA ATG TTT GCG AAA TTG GAA GCG AGC GCG TTA AAC GTC GAT GCA GAT GTC ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   M   F   A   K   L   E   A   S   A   L   N   V   D   A   D   V   M

AAT CAA GTA GTG AAT CCA CAT ACT GCA AAT TTA ATT CCA GCT GCG TTG TTA AAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   Q   V   V   N   P   H   T   A   N   L   I   P   A   A   L   L   K

CCA TTA CGC GGT ATC CTC TAT GCA GGT CTG CAT AAT GTT TAC TTA GTC GGT GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   L   R   G   I   L   Y   A   G   L   H   N   V   Y   L   V   G   A

GGC TTA GTT GTT GTC GCT CTT TTG TTA AAT ATT TTC GCA AAA GCG CAA CGA GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   L   V   V   V   A   L   L   L   N   I   F   A   K   A   Q   R   A

AAG GTT TAG 3'
--- --- ---
 K   V   *
```

Fig. 30B

```
5' ATG GCT GGT ATT GGC TTT AGC TTA GTG ATG CCA TTT ATG CCT CTT TAT ATC AAC
    M   A   G   I   G   F   S   L   V   M   P   F   M   P   L   Y   I   N

   ACA TTA GGT ACT TTT ACC CAC CAA CAA TTG AAT TTT TGG AGC GGA ATC ACG TTC
    T   L   G   T   F   T   H   Q   Q   L   N   F   W   S   G   I   T   F

   TCT TCC ACT TTT TTA GTC ACA ACC ATT GTT TCA CCT TGG TGG GGC CGT TTA GCC
    S   S   T   F   L   V   T   T   I   V   S   P   W   W   G   R   L   A

   GAC CGA AAA GGA CGC AAG TTA ATG TTA TTA CGC GCT TCA TTA GGG ATG GCT ATC
    D   R   K   G   R   K   L   M   L   L   R   A   S   L   G   M   A   I

   GTC ATT AGT TTA ATG GGC GCA GTC ACA AGT GTT TAT CAA TTG ATT GGC TTA CGC
    V   I   S   L   M   G   A   V   T   S   V   Y   Q   L   I   G   L   R

   TTA TTA CAA GGT GTT TTT TCT GGT TAT ATC AGT AAT GCG ACT GCG TTA GTT GCT
    L   L   Q   G   V   F   S   G   Y   I   S   N   A   T   A   L   V   A

   ACA GGC ACA CCT AAG GAA AAA AGC GGT CAA GTG CTT GGT ACT TTA GCC ACT GGT
    T   G   T   P   K   E   K   S   G   Q   V   L   G   T   L   A   T   G

   TCT GTT ACC GGC ACG TTG CTT GGT CCT CTT TTG GGC GGT GTC ACT GCG TCT ATT
    S   V   T   G   T   L   L   G   P   L   L   G   G   V   T   A   S   I

   TTC GGT TAT CGG CCA ACC TTT TTC ATT ACT GGT ACG ATT TTA TTA CTC GTT TTT
    F   G   Y   R   P   T   F   F   I   T   G   T   I   L   L   L   V   F

   GTT TTG AGT CTC GTC TTT GTC CAT GAA GAG TTT GTG CCA ATT GAA AAA AAT CAA
    V   L   S   L   V   F   V   H   E   E   F   V   P   I   E   K   N   Q

   GCC GCA TCT GGG AAA CAA ATT TTA AAA AAA CTA GAA CAT CCC CAC GTG ATT CTT
    A   A   S   G   K   Q   I   L   K   K   L   E   H   P   H   V   I   L

   GGA ATG TTT ATT ACG ACA TTA ATC ATT CAA GCT TCC AAT AAT TCA ATT AGT CCC
    G   M   F   I   T   T   L   I   I   Q   A   S   N   N   S   I   S   P

   ATC ATC AGT TTA TAT ATT CAA CAA TTG TTG GGT GGT CAC GGA AAT GTC ACC TTA
    I   I   S   L   Y   I   Q   Q   L   L   G   G   H   G   N   V   T   L

   ATT AGC GGA GTC ATT GCT TCT ATT CCA GGG ATT GCA ACG TTA ATA GCC GCC CCT
    I   S   G   V   I   A   S   I   P   G   I   A   T   L   I   A   A   P

   CGT TTC GGC CGG TTA GGC GAT CGT ATT GGC AGT GAA CGC ATC TTG ACA ATT GGC
    R   F   G   R   L   G   D   R   I   G   S   E   R   I   L   T   I   G

   TTA ATT TTA GCC ATT TTT GTT TAC CTA CCA ATG GCC TTT GTT CAA AAT GTC TGG
    L   I   L   A   I   F   V   Y   L   P   M   A   F   V   Q   N   V   W
```

Fig. 31A

```
CAG CTT GCC ATG CTT CGT TTC TTA GTT GGG ATT TCC GAT GCC TGT TTA CTA CCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   L   A   M   L   R   F   L   V   G   I   S   D   A   C   L   L   P

GCT GTT CAA ACG TTG ATC ACT CGT TAT TCC CCA AGT GAC GCT GCA GGT CGT ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   V   Q   T   L   I   T   R   Y   S   P   S   D   A   A   G   R   I

TTC AGT TAC AAT CAA TCT TTC CAA GCG ACA GGA AAT GTG ATT GGT CCT ATG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   S   Y   N   Q   S   F   Q   A   T   G   N   V   I   G   P   M   I

GGT TCA AGC GTA TCT GCC GCT TTC GGT TAT CGT GGG GTC TTT ATT TCC ACT TCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   S   S   V   S   A   A   F   G   Y   R   G   V   F   I   S   T   S

TGT CTG GTT CTC CTT AAC CTT CTA TGG GTT CGT CGA AGC ACC GCC GAA TTA AAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 C   L   V   L   L   N   L   L   W   V   R   R   S   T   A   E   L   K

AAA GAG AAA AAT GAT GAC TAA 3'
--- --- --- --- --- --- ---
 K   E   K   N   D   D   *
```

Fig. 31B

```
5'  ATG CAA TTT AAT TTG AAA CTT ACA GAA AAA GGA GCG ATA TAT GTG ACT TCA CAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   Q   F   N   L   K   L   T   E   K   G   A   I   Y   V   T   S   Q

    CAA CCT GTG GAT ATT CAC GGA AAA CCT TAT AAC CGG TCG TTA CTA GTC GGT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   P   V   D   I   H   G   K   P   Y   N   R   S   L   L   V   G   V

    TTA CTA ATC GGA ACA TTT TGT ACG ATT TTA AAC CAA ACG CTT TTA ACG ACG GCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   L   I   G   T   F   C   T   I   L   N   Q   T   L   L   T   T   A

    TTA CCT ACA TTA ATG AAA GAA TTT GAT ATC TCA GCT TCA AGT GTA CAG TGG CTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   P   T   L   M   K   E   F   D   I   S   A   S   S   V   Q   W   L

    ACC ACT GGA TTC CTA CTT GTG AAC GGC ATT ATG ATT CCC ATC AGT GCT TGG TTG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   T   G   F   L   L   V   N   G   I   M   I   P   I   S   A   W   L

    ATT AAC AAA TTC AGT TCA AAA AAA TTA TAC ATA ACA GCG ATG TCA ACG TTC TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   N   K   F   S   S   K   K   L   Y   I   T   A   M   S   T   F   L

    ATC GGA ACC ATC ATC TGT TTT GTA GCG CAA GAC TTC GGC ATG TTA TTG ACT GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   G   T   I   I   C   F   V   A   Q   D   F   G   M   L   L   T   G

    CGT CTT GTT CAA GCC GCA GGG GTC GGT GTA TCT ATG CCT CTT TTA CAA ACA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   L   V.  Q   A   A   G   V   G   V   S   M   P   L   L   Q   T   I

    ATG CTT TCA ATT TTC CCG CCT GAA AAA CGT GGC GCT GCC ATG GGC ACA ACG GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   L   S   I   F   P   P   E   K   R   G   A   A   M   G   T   T   G

    ATT GTT ATC GGA TTA GCA CCC GCT TTA GGT CCA ACA TTA TCT GGT TGG ATT ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   V   I   G   L   A   P   A   L   G   P   T   L   S   G   W   I   I

    GAT TCT TAT ACA TGG CGT GAC CTT TTT GGG ATG GTC ATT CCA ATT GTC GTT TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     D   S   Y·  T   W   R   D   L   F   G   M   V   I   P   I   V   V   L

    GTC CTA ATT TTA GCT TCA TTT TTA ATG AAA AAT GTC ATT CAA TTA TCT AAT CCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   L   I   L   A   S   F   L   M   K   N   V   I   Q   L   S   N   P

    AGT ATC GAT GTC TTA TCG GTC ATT CTT TCT ACT CTC GGC TTT GGT AGC TTA CTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   I   D   V   L   S   V   I   L   S   T   L   G   F   G   S   L   L

    TAC GGT TTC TCA AGC GTT GGT GAT AAA GGT TGG GGT AGC CCG CAA GTC TAT GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Y   G   F   S   S   V   G   D   K   G   W   G   S   P   Q   V   Y   G

    TTC TTA ATT GTT GGT GCG ATT GTT TTA TGT CTT TTT ACA TAT CGA CAA TTG CAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   L   I   V   G   A   I   V   L   C   L   F   T   Y   R   Q   L   H

    TTA GAA CAA CCT TTC TTA GAA CTA CGC GTT TTT AAA TCA AAA GTC TTT ACT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   E   Q   P   F   L   E   L   R   V   F   K   S   K   V   F   T   V
```

Fig. 32A

```
GCG GCC ATT TTA TCT GGC GTA ACC AAC ATG GCA ATG ATT GGC GCC GAA ATG GTT
 A   A   I   L   S   G   V   T   N   M   A   M   I   G   A   E   M   V

TTA CCT TTG TAT ATC CAA AAT ATT CGT GGC GAA TCT GCC TTC CAT TCA GGG TTA
 L   P   L   Y   I   Q   N   I   R   G   E   S   A   F   H   S   G   L

ATG CTT TTA CCA GGA GCG TTA GTC ATG GGA CTT ATG ATG CCA GTA ACT GGT CGC
 M   L   L   P   G   A   L   V   M   G   L   M   M   P   V   T   G   R

ATT TTT GAT AAA ATT GGT GCC CGT CGT CTT GCG ATT ACT GGG ATG TTT ATT TTA
 I   F   D   K   I   G   A   R   R   L   A   I   T   G   M   F   I   L

ACA GCT GCA ACG TTA CCG TTT GCT TTC TTA ACA AAA GCA ACA CCT ATT ATT TAT
 T   A   A   T   L   P   F   A   F   L   T   K   A   T   P   I   I   Y

ATC ATT GTC CTA TAT GCG ATT CGG ATG TTT GGT ATT TCG ATG GTC ATG ATG CCT
 I   I   V   L   Y   A   I   R   M   F   G   I   S   M   V   M   M   P

GTG ACA ACT TCT GGC ATG AAT GCT TTA CCA ATG AAT TTA TTA AGT CAC GGG ACT
 V   T   T   S   G   M   N   A   L   P   M   N   L   L   S   H   G   T

GCT GTA AAC AAC ACC TTC CGA CAA GTA GCC AGC TCA ATT GGA ACA GCT GTT TTG
 A   V   N   N   T   F   R   Q   V   A   S   S   I   G   T   A   V   L

ATT AGT GTT TTA ACA AAT GTT ACC AAA GAC GGC TTA CCT GCA TCG GAC TTA TTA
 I   S   V   L   T   N   V   T   K   D   G   L   P   A   S   D   L   L

AAA ACA GCC CCA CTC ACT TAT CGC GAT CAA GCA ACA AAC GCA ACA CTA AAC GGT
 K   T   A   P   L   T   Y   R   D   Q   A   T   N   A   T   L   N   G

TAT CAC GCA GCC TTC TTC GTG GCA ACA ATC TTT GGC GTT CTC GGC TTG GCG ATT
 Y   H   A   A   F   F   V   A   T   I   F   G   V   L   G   L   A   I

ACT TTC TTC TTA AAT AAA AAA GAA GCG ATG CCT GTT AAG GAA GTA GGT GCA ATG
 T   F   F   L   N   K   K   E   A   M   P   V   K   E   V   G   A   M

AAA TAA 3'
 K   *
```

Fig. 32B

```
5'  ATG GAG AAA ATT AAA AGT TAC CAA GAG GAT TCT GAA GTG CAG AAA AAT CGT TGG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   E   K   I   K   S   Y   Q   E   D   S   E   V   Q   K   N   R   W

    TGG ATC TTA GTT TCT GTG GCT ATG TTT ACC TTC ATG TCT ACA TTA GAT TCC AGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     W   I   L   V   S   V   A   M   F   T   F   M   S   T   L   D   S   S

    ATT GTA AAT ATT GCT TTA CCG ACT ATT TCA AAA GAA ATG TCC GTA CCC ATG AAT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   V   N   I   A   L   P   T   I   S   K   E   M   S   V   P   M   N

    CAA GCG GAA TGG GTC GTA TCT ATT TAT TTA ATG GTT GTC TGT GCT TGT TTA TTA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   A   E   W   V   V   S   I   Y   L   M   V   V   C   A   C   L   L

    TTA TTT GGA AAA ATT GGC GAC AGT TTT GGA AAA ATA AAA GTT TAT CGA ATT GGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   F   G   K   I   G   D   S   F   G   K   I   K   V   Y   R   I   G

    ACG GTT ATT TTT ACG ATT GGT TCG TTA CTT TGT GGC TTT AAT CAA TCG CTG AGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   V   I   F   T   I   G   S   L   L   C   G   F   N   Q   S   L   S

    TTT CTA TTA TTT GCG CGA GTC GTT CAA AGT ATT GGT GCT AGC ATG ACC ATG GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   L   L   F   A   R   V   V   Q   S   I   G   A   S   M   T   M   A

    ACC AAT TCA GGA ATT ATT ACG GAA GTA TTC CCT TTA AAT GAA AGA GGG CGG GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   N   S   G   I   I   T   E   V   F   P   L   N   E   R   G   R   A

    CTT GGT GCA GTA GGT GCC TTT GTC TCT TTA RGC GCG ATT GCT GGT CCT GGG ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   G   A   V   G   A   F   V   S   L   X   A   I   A   G   P   G   I

    GGT GGA TTG ATC TTA TCA AAC TTT TCT TGG TCG TAT ATT TTT TGG ATT AAT GTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   G   L   I   L   S   N   F   S   W   S   Y   I   F   W   I   N   V

    CCC GTG GGG CTA GTT ACA ATT TTA ATT GGC GAA AAG TTC CTA CCA AAA GAT ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     P   V   G   L   V   T   I   L   I   G   E   K   F   L   P   K   D   I

    ACA AAA ACA AAA GAA AAA ATC GAT TTT TCA GGA TTT GCT TGT ATT GCC ATT GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   K   T   K   E   K   I   D   F   S   G   F   A   C   I   A   I   A

    ATT ATG ACG TTC TTC GGC GGG ATT TTC CTC GGA CAA GAG AGC GGG TTT GGC TCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   M   T   F   F   G   G   I   F   L   G   Q   E   S   G   F   G   S

    TTA CAA TCC TAT TTA TTG TTC ATT ATT GCA GTA ATC GCA TTA GGA TTA TTT ATA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   Q   S   Y   L   L   F   I   I   A   V   I   A   L   G   L   F   I

    ATG GTT GAA CGA AAA AGA AAA AGC CCA CTT ATT AAG TTC GCA ATT TTT AAA AAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   V   E   R   K   R   K   S   P   L   I   K   F   A   I   F   K   N

    AAA ATT TTT ACT TTA AGT TTA TTG TCC GCC GTT TTA ATT TTT GCC TCA AAC TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     K   I   F   T   L   S   L   L   S   A   V   L   I   F   A   S   N   F
```

Fig. 33A

```
TTT GTA AAT GTT GTC ATT CCA TTT TAT TTG CAG GAT GCA CGC AAA CTA AGT GCT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 F    V    N    V    V    I    P    F    Y    L    Q    D    A    R    K    L    S    A

AGT TAT GCT GGT CTA TTA ATG ATG GTA TTT CCG TTA TTA ATG GTG GTT GGG GCT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 S    Y    A    G    L    L    M    M    V    F    P    L    L    M    V    V    G    A

CCT CTG AGT GGC TAT TTG ACG GAT AAA ATT GGC CCA GGT ATT TTA ACA TTT GGC
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 P    L    S    G    Y    L    T    D    K    I    G    P    G    I    L    T    F    G

GGA TTG TTG CTC TTG TGC TGT ACG TCG TTA ATG TAT ATG TTT TTA GAT ATG AAT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 G    L    L    L    L    C    C    T    S    L    M    Y    M    F    L    D    M    N

TCG CCT ATC TGG TAT TAT GTG ATT GCA ACG GCC ATT ATG GGC TTG GGA AAT GCA
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 S    P    I    W    Y    Y    V    I    A    T    A    I    M    G    L    G    N    A

CTT TTC CAG TCT CCA AAC AAT ACA ATG GTT ATG AGC AGT GTT GAA AAG CAA GAT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 L    F    Q    S    P    N    N    T    M    V    M    S    S    V    E    K    Q    D

TTA GGT GTA GCA GGA AGT ATG AAT TCT TTT GCT AGA AAC TTA GGA ATG GTC ATT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 L    G    V    A    G    S    M    N    S    F    A    R    N    L    G    M    V    I

GGG ATT GCG TTG TCA ACG ACC ATT TTA TAT CGT GGC ATG AGC GAA GCC TAT GGT
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 G    I    A    L    S    T    T    I    L    Y    R    G    M    S    E    A    Y    G

GAA CGA GTA ACC ACG TAT CTG GCT AAT CGC CCA GAT ATA TTT ATT GTG GGA ATG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 E    R    V    T    T    Y    L    A    N    R    P    D    I    F    I    V    G    M

CGT GAA ACC TTT TTT GTC GCC TTT CTA TTA TGT GTG GCG GCC TTT ATA TTA ACG
---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 R    E    T    F    F    V    A    F    L    L    C    V    A    A    F    I    L    T

ATT TTA CGT TTT CGG AAA ACA ACC AAA TAA 3'
---  ---  ---  ---  ---  ---  ---  ---  ---  ---
 I    L    R    F    R    K    T    T    K    *
```

Fig. 33B

```
5'  ATG GAA CAA AAA AAC GTA CGA TTA TTC CCT GCC GTT TTA GCG ACA GGA ATT ATG
    M   E   Q   K   N   V   R   L   F   P   A   V   L   A   T   G   I   M

    TCT TTT GCT GGC GTA TTA ATT GAA ACA GCA ATG AAT GTA ACC TTT CCG ACA TTA
    S   F   A   G   V   L   I   E   T   A   M   N   V   T   F   P   T   L

    ACA AAG GAA TTT GGT GTT TCC ACC GGT ACC GTG CAG TGG GTG ACA ACG ATT TAT
    T   K   E   F   G   V   S   T   G   T   V   Q   W   V   T   T   I   Y

    TTA TTA GTT ATT TCT ATC ATG GTG CCC TTA TCC AAT TAT TTA TTA AAA ACC TAT
    L   L   V   I   S   I   M   V   P   L   S   N   Y   L   L   K   T   Y

    TCG TTA AGA CGC TTA TTT ATT GTT GCG AAT CTT TTC TTT TTG ATT GGT TTA GCG
    S   L   R   R   L   F   I   V   A   N   L   F   F   L   I   G   L   A

    ATT GAT GTC TAT TCG CCA TCT TTC AGT ATC TTG TTA TTA GGC CGA CTC TTC CAA
    I   D   V   Y   S   P   S   F   S   I   L   L   L   G   R   L   F   Q

    GGG GCT AGC ACT GGG ATT GCT TTG CCA CTG ATG TTT CAT ATT ATT TTG AAC TTT
    G   A   S   T   G   I   A   L   P   L   M   F   H   I   I   L   N   F

    ACC CCG CTG GAA AAA CGA KGA ACG ATG ATG GGG GTA SGC ACA TTG ACC ACT TCA
    T   P   L   E   K   R   X   T   M   M   G   V   X   T   L   T   T   S

    ATT GCT CCA GCC ATT GGT CCA ACG TAT GGT GGA ATC TTA ACA TCT TCC CTG TCT
    I   A   P   A   I   G   P   T   Y   G   G   I   L   T   S   S   L   S

    TGG CAC GCC ATC TTT TTA TTC TTA ATT CCT ATT TTA CTG CTT TCT TTG TTT ATG
    W   H   A   I   F   L   F   L   I   P   I   L   L   L   S   L   F   M

    GGC TTA TCT GCT ATT CCA GAA ATA CCT GTT AAA AAA ACC ACC ACT CTT GAT TTG
    G   L   S   A   I   P   E   I   P   V   K   K   T   T   T   L   D   L

    GTT AGT TTA ATT GGT ATT GCC CTG TTA TTT AGT GGT TTG TTG ATG TTT TTA AGT
    V   S   L   I   G   I   A   L   L   F   S   G   L   L   M   F   L   S

    AAA ATT GGC ACA TTG TTT GGC TGG CTT TCA CTT CTG GCA GCT GTT ATT GGC TTT
    K   I   G   T   L   F   G   W   L   S   L   L   A   A   V   I   G   F

    GTT ATT TTT TAT AAA CGA GCA ACG ACC GCT GAA CAG CCA CTA GTT CGT TTG ACA
    V   I   F   Y   K   R   A   T   T   A   E   Q   P   L   V   R   L   T

    ATT TTA AAA AAT CCC GCC TTC GTT TTA TTT TTA TGT GGC TTT TTA GTT TGT CAA
    I   L   K   N   P   A   F   V   L   F   L   C   G   F   L   V   C   Q

    TTC TTA TTG TTA GGT ATT TCC TTT GTC TTA CCA AAT TTT GTT CAA ATC GTC CTT
    F   L   L   L   G   I   S   F   V   L   P   N   F   V   Q   I   V   L
```

Fig. 34A

```
GGA AAA AAT GCG TTT GTA GCT GGG CTA GTC ATG CTG CCA GGT GCT ACA GTT GGG
 G   K   N   A   F   V   A   G   L   V   M   L   P   G   A   T   V   G

GCC ATT TTG GCA CCA CTT TCT GGA CGA GTG TTG GAT CAA TAC GGC GCT AAA AAA
 A   I   L   A   P   L   S   G   R   V   L   D   Q   Y   G   A   K   K

CCT ATA TTA TTT GGC TTA AGT TTG GCG ACA ATC GGT TGG CTA GCT TTA ACA ATT
 P   I   L   F   G   L   S   L   A   T   I   G   W   L   A   L   T   I

TTA CTA GAG ATG CCT GTT TTA TTA GGA TTC GTC GCT GGA CAT GTT ACC TAT ATG
 L   L   E   M   P   V   L   L   G   F   V   A   G   H   V   T   Y   M

ATT GGT CTA GGT TTT GCA TAC AGT AAT ATG ATG ACG ACT GGT ATG AGC TTG TTA
 I   G   L   G   F   A   Y   S   N   M   M   T   T   G   M   S   L   L

GAG GAG AAA GAT TTT GGC GAC GGC AAT ACG TTA TTT AAT ACA CTT CAA CAA TTT
 E   E   K   D   F   G   D   G   N   T   L   F   N   T   L   Q   Q   F

TCT GGC GCT ATT GCG ACG GCC ATT GTA GCG ACG ATT ATT AAT ATC GCT CAA GAC
 S   G   A   I   A   T   A   I   V   A   T   I   I   N   I   A   Q   D

CAC GCA GAT AAT TTT GCT CAG GGG ACA ACG ATG GGC TCC CTC ATT TCT CTG ATT
 H   A   D   N   F   A   Q   G   T   T   M   G   S   L   I   S   L   I

TTC TTG TTA ACC TTA TTA GTA ATT GTT TTA ATT GCT TGC TGG AAT TAT TTC CGC
 F   L   L   T   L   L   V   I   V   L   I   A   C   W   N   Y   F   R

AAA AAA GCT TAA 3'
 K   K   A   *
```

Fig. 34B

```
5' ATG AAA AAC ACC CGA AAC ACC GTC CCA TCA TTA CTA ATT ATT ATT TTG CTT GTG
   M   K   N   T   R   N   T   V   P   S   L   L   I   I   I   L   L   V

   GGA TTT CCA CAA ATC AGC GAG TCG ATT TTC ACA CCC ATT TTG CCA GCG CTT AGT
   G   F   P   Q   I   S   E   S   I   F   T   P   I   L   P   A   L   S

   TCT GCT TTT AAC ATT CAC GAT AGC GTG GCA CAA TTA ACC ATG AGT ATT TAT TTT
   S   A   F   N   I   H   D   S   V   A   Q   L   T   M   S   I   Y   F

   GTT GCC TTT GCC GTT GGC GTT TTA TTC TGG GGA ATT CTA TCC GAT ACG ATT GGA
   V   A   F   A   V   G   V   L   F   W   G   I   L   S   D   T   I   G

   CGT CGC CCA GCT ATG TTA TTA GGA ATT TTA TGC TAT TTA ATT GGA AAT ATT GGC
   R   R   P   A   M   L   L   G   I   L   C   Y   L   I   G   N   I   G

   CTT TTT TTG AGT CCG ACA TTT CCT TGG CTG TTA TTC TTC CGT TTA ATT CAA GCA
   L   F   L   S   P   T   F   P   W   L   L   F   F   R   L   I   Q   A

   TTT GGT GCA AGC GTT GGT TCT GTC ATT ACC CAA TCG ATG ATG CGG GAA AGC TTT
   F   G   A   S   V   G   S   V   I   T   Q   S   M   M   R   E   S   F

   AAC GAT CTT CAA AGA GGG AAA ATT TTT GCT CAA ACA AGC GCT GCC CTA TCG CTA
   N   D   L   Q   R   G   K   I   F   A   Q   T   S   A   A   L   S   L

   TCT CCT GCC ATC GGA CCC TTA GTC GGC GGA CTT GTT CAA ACC TAT TTT GGC TAT
   S   P   A   I   G   P   L   V   G   G   L   V   Q   T   Y   F   G   Y

   CAG CAT GTC TTT AGC ACC TTA ATT GTC ATG GCT GTT CTA TTA CTT TTC TAT AGT
   Q   H   V   F   S   T   L   I   V   M   A   V   L   L   L   F   Y   S

   TTT TAT CGT TTA CCA GAA ACC TTT CAC TTA GGA AAT ACC TTA ACG AAA GCG CCT
   F   Y   R   L   P   E   T   F   H   L   G   N   T   L   T   K   A   P

   TTT TGG TCC GTG GCG AAA CGC CTT TCC TTC GAC TCG AAA GTT TGG TGC TAC GGC
   F   W   S   V   A   K   R   L   S   F   D   S   K   V   W   C   Y   G

   TTA TTG ATT GGG GGC ATT AAT GGT TTA TTG TTT AGC TAC TAT GCA GAA GCT CCG
   L   L   I   G   G   I   N   G   L   L   F   S   Y   Y   A   E   A   P

   TTT ATT TTC ATC AAT CAC TTT GGC TTT AGC AGT GTC CAA TAT GGC TTT CTT GGT
   F   I   F   I   N   H   F   G   F   S   S   V   Q   Y   G   F   L   G

   TTA GTT TTA GCT GTG TCT AGT ATT CTT GCT GCG GTT CTT ATC AAT TGG TTG CTA
   L   V   L   A   V   S   S   I   L   A   A   V   L   I   N   W   L   L

   GGA TAT TGG TCA GCG GAA AAA ATT GCC TTT GCT GGT TTG CTG CTA AGT ATC TTT
   G   Y   W   S   A   E   K   I   A   F   A   G   L   L   L   S   I   F
```

Fig. 35A

```
GCG GCT TGC TGT TTA TTA ATA ATG ACA CAA TTT GAA ACA ATC ATT GGT CTT ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   A   C   C   L   L   I   M   T   Q   F   E   T   I   I   G   L   I

TTG TTT ATT TTT CTC TTT TTC TTA GGG ATT AAT ATT ACT TTG CCC AAT GCC TTA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   F   I   F   L   F   F   L   G   I   N   I   T   L   P   N   A   L

AGT ATG GCC TTA AAA GGC TAT GAA TCT GTA ATT GGT ACT GCC AGT GGG ATT TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   M   A   L   K   G   Y   E   S   V   I   G   T   A   S   G   I   F

AGT TTT GTC TAT TAC TTG TTT GTT AGT TTC TTT ACC TAT TTG ATT AGT TAC TTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   F   V   Y   Y   L   F   V   S   F   F   T   Y   L   I   S   Y   F

CAC AAT GGT ACC ATT TGG GTC TTG CCG CTT TAC TTT TTA AGT TTT GCT TGT CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 H   N   G   T   I   W   V   L   P   L   Y   F   L   S   F   A   C   L

TTA GCT TTC AGC TAT TAT GTC ATT GTT TAT CGT AAA ACT TTA AAA TAA 3'
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   A   F   S   Y   Y   V   I   V   Y   R   K   T   L   K   *
```

Fig. 35B

```
5'  GTG GGG CTT AAC ATT GTC TTA CCA CTC GTT AAT ATT ATT AAT GGA TTA GGT TGG
    V  G  L  N  I  V  L  P  L  V  N  I  I  N  G  L  G  W

    ATG TTT GGC GTT GGC GGC GCT ACG TTG TTT TCG ACC ACG GTT GCG CAA AAA GAA
    M  F  G  V  G  G  A  T  L  F  S  T  T  V  A  Q  K  E

    ATC AAA AAA GCC AAT CAG TAT TTT AGT TTA ACG ATT GGG TTA GTT TTC GTT ATA
    I  K  K  A  N  Q  Y  F  S  L  T  I  G  L  V  F  V  I

    GGG AGT CTC TTT ACT TTA GTG AGC TTG ATT TTT TCC GAT CAA ATT ATT CGG GGG
    G  S  L  F  T  L  V  S  L  I  F  S  D  Q  I  I  R  G

    CTT CAG GGG ACC GGT GTC TTG TTT GGT TTG GCA AAG GAA TAT TAT ATG ATT TAC
    L  Q  G  T  G  V  L  F  G  L  A  K  E  Y  Y  M  I  Y

    CTT AGC TGT TCA TTA TTA TTT ATT TTA AAC AAT TGC CTG ATT ACT TTT TTA AGA
    L  S  C  S  L  L  F  I  L  N  N  C  L  I  T  F  L  R

    AAT GAC CAT AAT CCT CGT TTG GCA ACG ATT GCT TTT GTC AGT GGG GGA ATT GTC
    N  D  H  N  P  R  L  A  T  I  A  F  V  S  G  G  I  V

    AAC ATT ATT TTG GAT TAT GTC TTT ATT TAT CAA TTT GGG TGG GGT ATG GCA GGT
    N  I  I  L  D  Y  V  F  I  Y  Q  F  G  W  G  M  A  G

    GCG GCC ATT GCA ACG GTG ATG TCA CCG TTA ACA AGC TTA ATC CTG ATT ACG TTG
    A  A  I  A  T  V  M  S  P  L  T  S  L  I  L  I  T  L

    CAT AAG TGG TCG CCG CAA CGA GTT TTA CGT TTT GAA AAG TTT AAA GTG AAA TTT
    H  K  W  S  P  Q  R  V  L  R  F  E  K  F  K  V  K  F

    CAG GAT GTT CGA GAA ATT ATG TCG ATT GGT TTT TCC TCC TTT TTA AAT GAA TTT
    Q  D  V  R  E  I  M  S  I  G  F  S  S  F  L  N  E  F

    TCT TCG GCC TTT GTC ATG TTT TTA TTT AAT ATT GTT TTG TTA AAC TTA GTT GGT
    S  S  A  F  V  M  F  L  F  N  I  V  L  L  N  L  V  G

    CAT GTG GCG ATT TCA GCT TAT GCG ATT GTT GCC AAT CTC AAT ATT ATT GTG ATT
    H  V  A  I  S  A  Y  A  I  V  A  N  L  N  I  I  V  I

    GCT ATT TTT ACA GGG ATT GGG CAA GGG GCG CAA CCA TTA CTC AGC CGA TAC TAT
    A  I  F  T  G  I  G  Q  G  A  Q  P  L  L  S  R  Y  Y

    GGT TTA GGG GAA ACG AAG GTG TTG CGT AAA TTT GTT AAA CTT AGC TTT ATT ACG
    G  L  G  E  T  K  V  L  R  K  F  V  K  L  S  F  I  T

    TAC TTG GTC GCC GGC TTC CTG TTC TTT TTA ATC AGT CAA GTG TTT ACG GGA CAA
    Y  L  V  A  G  F  L  F  F  L  I  S  Q  V  F  T  G  Q
```

Fig. 36A

```
ATT ATT GAG GTC TTT AAT AGC GAA GGG AAC GAC CAG TTG GCG CAA ATT GCT AGG
 I   I   E   V   F   N   S   E   G   N   D   Q   L   A   Q   I   A   R

ACA GCC ATC CGC TTA TAT GCG ATT TCC TTC TTG TTT ACA GGA TTG AAT TTT ATG
 T   A   I   R   L   Y   A   I   S   F   L   F   T   G   L   N   F   M

GGG ATT TAT TAC TTT TCG GCA GTT CGT AAA CCG AAA ATG GCG CTA ATG ATT TCC
 G   I   Y   Y   F   S   A   V   R   K   P   K   M   A   L   M   I   S

TCT TTG CGT GGG TTC TTT TTA ATT GTG CCA GTT TTA TTT ATT TTG GTG AAA TTA
 S   L   R   G   F   F   L   I   V   P   V   L   F   I   L   V   K   L

CTA GGA TTA ACC GGT GTT TGG TTA GCC ATG CCA GTA GTT GAA TTT GTT ACG TTT
 L   G   L   T   G   V   W   L   A   M   P   V   V   E   F   V   T   F

GGA CTA ATG CTT GTG GGC TAC CTT GCG TAT CGA AAC TAT TTG AAA AAA AGA GAA
 G   L   M   L   V   G   Y   L   A   Y   R   N   Y   L   K   K   R   E

GCA GTA ACC TGA 3'
 A   V   T   *
```

Fig. 36B

```
5'  ATG CCA TAT TTA TAT TTG TTA TTA GCA ATA GTC AGT GAA GTA ATA GGC AGT GGA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   P   Y   L   Y   L   L   L   A   I   V   S   E   V   I   G   S   A

    TTT TTA AAA TCT TCA GAC GGT TTC TCA AAA TTA TAT CCA ACT ATA ACA ACA ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   L   K   S   S   D   G   F   S   K   L   Y   P   T   I   T   T   I

    ATT TCA TAC CTT ATT TCT TTT TAT TTT CTG AGT AAA ACT ATG CAA CAT TTA CCA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   S   Y   L   I   S   F   Y   F   L   S   K   T   M   Q   H   L   P

    CTT AAT ATT GCT TAC GCA AGT TGG TCA GGT TTA GGA TTA GTA TTA ACA ACA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   N   I   A   Y   A   S   W   S   G   L   G   L   V   L   T   T   I

    GTT TCA GTT CTT ATT TTC AAA GAA CAA ATA AAT TTA ATA AGC ATT ATT TCA ATT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   S   V   L   I   F   K   E   Q   I   N   L   I   S   I   I   S   I

    ATT TTA ATA ATA TTT GGT GTT GTT TTA CTA AAC ACA TTC GGA TCA TCA CAC TAA 3'
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   L   I   I   F   G   V   V   L   L   N   T   F   G   S   S   H   *
```

Fig. 37